Health Technology Assessment 2006; Vol. 10: No. 1

Appendices

Go to main text

The clinical and cost-effectiveness of donepezil, rivastigmine, galantamine and memantine for Alzheimer's disease

E Loveman, C Green, J Kirby, A Takeda, J Picot, E Payne and A Clegg

January 2006

Health Technology Assessment
NHS R&D HTA Programme

How to obtain copies of this and other HTA Programme reports.
An electronic version of this publication, in Adobe Acrobat format, is available for downloading free of charge for personal use from the HTA website (http://www.hta.ac.uk). A fully searchable CD-ROM is also available (see below).

Printed copies of HTA monographs cost £20 each (post and packing free in the UK) to both public **and** private sector purchasers from our Despatch Agents.

Non-UK purchasers will have to pay a small fee for post and packing. For European countries the cost is £2 per monograph and for the rest of the world £3 per monograph.

You can order HTA monographs from our Despatch Agents:

– fax (with **credit card** or **official purchase order**)
– post (with **credit card** or **official purchase order** or **cheque**)
– phone during office hours (**credit card** only).

Additionally the HTA website allows you **either** to pay securely by credit card **or** to print out your order and then post or fax it.

Contact details are as follows:

HTA Despatch
c/o Direct Mail Works Ltd
4 Oakwood Business Centre
Downley, HAVANT PO9 2NP, UK

Email: orders@hta.ac.uk
Tel: 02392 492 000
Fax: 02392 478 555
Fax from outside the UK: +44 2392 478 555

NHS libraries can subscribe free of charge. Public libraries can subscribe at a very reduced cost of £100 for each volume (normally comprising 30–40 titles). The commercial subscription rate is £300 per volume. Please see our website for details. Subscriptions can only be purchased for the current or forthcoming volume.

Payment methods

Paying by cheque
If you pay by cheque, the cheque must be in **pounds sterling**, made payable to *Direct Mail Works Ltd* and drawn on a bank with a UK address.

Paying by credit card
The following cards are accepted by phone, fax, post or via the website ordering pages: Delta, Eurocard, Mastercard, Solo, Switch and Visa. We advise against sending credit card details in a plain email.

Paying by official purchase order
You can post or fax these, but they must be from public bodies (i.e. NHS or universities) within the UK. We cannot at present accept purchase orders from commercial companies or from outside the UK.

How do I get a copy of *HTA on CD*?

Please use the form on the HTA website (www.hta.ac.uk/htacd.htm). Or contact Direct Mail Works (see contact details above) by email, post, fax or phone. *HTA on CD* is currently free of charge worldwide.

The website also provides information about the HTA Programme and lists the membership of the various committees.

Appendix 1

Diagnostic criteria

DSM-IV criteria for the diagnosis of dementia of the Alzheimer's type

A. The development of multiple cognitive deficits manifested by both:

1. Memory impairment (impaired ability to learn new information or to recall previously learned information).
2. One or more of the following cognitive disturbances:
   (a) aphasia (language disturbance)
   (b) apraxia (impaired ability to carry out motor activities despite intact motor function)
   (c) agnosia (failure to recognise or identify objects despite intact sensory function)
   (d) disturbance in executive functioning (i.e. planning, organising, sequencing, abstracting).

B. The cognitive deficits in criteria A1 and A2 each cause significant impairment in social or occupational functioning and represent a significant decline from a previous level of functioning.

C. The course is characterised by gradual onset and continuing cognitive decline.

D. The cognitive deficits in criteria A1 and A2 are not due to any of the following:

1. Other central nervous system conditions that cause progressive deficits in memory and cognition (e.g. cerebrovascular disease, Parkinson's disease, Huntington's disease, subdural hematoma, normal-pressure hydrocephalus, brain tumour).
2. Systemic conditions that are known to cause dementia (e.g. hypothyroidism, vitamin B or folic acid deficiency, niacin deficiency, hypercalcaemia, neurosyphilis, HIV infection).
3. Substance-induced conditions.

E. The deficits do not occur exclusively during the course of a delirium.

F. The disturbance is not better accounted for by another Axis I disorder (e.g. major depressive episode, schizophrenia).

ICD-10 criteria for the diagnosis of dementia in Alzheimer's disease

F00 Alzheimer's disease

A primary degenerative cerebral disease of unknown aetiology, with characteristic neuropathological and neurochemical features. It is usually insidious in onset and develops slowly but steadily over a period of years. This period can be as short as 2 or 3 years, but can occasionally be considerably longer. The onset can be in middle adult life or even earlier (Alzheimer's disease of presenile onset), but the incidence is higher in later life (Alzheimer's disease of senile onset). In cases with onset before the age of 65–70 years, there is the likelihood of a family history of a similar dementia, a more rapid course, and prominence of features of temporal and parietal lobe damage, including dysphasia or dyspraxia. In cases with a later onset, the course tends to be slower and to be characterised by more general impairment of higher cortical functions. People with Down's syndrome are at high risk of developing Alzheimer's disease.

There are characteristic changes in the brain: a marked reduction in the population of neurons, particularly in the hippocampus, substantia innominata, locus ceruleus, and temporoparietal and frontal cortex; appearance of neurofibrillary tangles made of paired helical filaments; neuritic (argentophil) plaques, which consist largely of amyloid and show a definite progression in their development (although plaques without amyloid are also known to exist); and granulovacuolar bodies. Neurochemical changes have also been found, including a marked reduction in the enzyme choline acetyltransferase, in acetylcholine itself, and in other neurotransmitters and neuromodulators.

As originally described, the clinical features are accompanied by the above brain changes. However, it now appears that the two do not always progress in parallel: one may be indisputably present with only minimal evidence of the other. Nevertheless, the clinical features of Alzheimer's disease are such that it is often possible to make a presumptive diagnosis on clinical grounds alone. Dementia in Alzheimer's disease is irreversible at present.

Diagnostic guidelines

The following features are essential for a definite diagnosis:

(a) Presence of a dementia as described above.
(b) Insidious onset with slow deterioration. While the onset usually seems difficult to pinpoint in time, realisation by others that the defects exist may come suddenly. An apparent plateau may occur in the progression.
(c) Absence of clinical evidence, or findings from special investigations, to suggest that the mental state may be due to other systemic or brain disease which can induce a dementia (e.g. hypothyroidism, hypercalcaemia, vitamin B_{12} deficiency, niacin deficiency, neurosyphilis, normal pressure hydrocephalus, or subdural haematoma).
(d) Absence of a sudden, apoplectic onset, or of neurological signs of focal damage such as hemiparesis, sensory loss, visual field defects, and incoordination occurring early in the illness (although these phenomena may be superimposed later).

In a certain proportion of cases, the features of Alzheimer's disease and vascular dementia may both be present. In such cases, double diagnosis (and coding) should be made. When the vascular dementia precedes the Alzheimer's disease, it may be impossible to diagnose the latter on clinical grounds.

Differential diagnosis

Consider: a depressive disorder (F30–F39); delirium (F05); organic amnesic syndrome (F04); other primary dementias, such as in Pick's, Creuzfeldt–Jakob or Huntington's disease (F02.–); secondary dementias associated with a variety of physical diseases, toxic states, etc. (F02.8); mild, moderate or severe mental retardation (F70–F72).

Dementia in Alzheimer's disease may coexist with vascular dementia (to be coded F00.2), as when cerebrovascular episodes (multi-infarct phenomena) are superimposed on a clinical picture and history suggesting Alzheimer's disease. Such episodes may result in sudden exacerbations of the manifestations of dementia. According to postmortem findings, both types may coexist in as many as 10–15% of all dementia cases.

F00.0 Dementia in Alzheimer's disease with early onset

Dementia in Alzheimer's disease beginning before the age of 65 years. There is relatively rapid deterioration, with marked multiple disorders of the higher cortical functions. Aphasia, agraphia, alexia, and apraxia occur relatively early in the course of the dementia in most cases.

Diagnostic guidelines

As for dementia, described above, with onset before the age of 65 years, and usually with rapid progression of symptoms. Family history of Alzheimer's disease is a contributory but not necessary factor for the diagnosis, as is a family history of Down's syndrome or of Iymphoma.

F00.1 Dementia in Alzheimer's disease with late onset

Dementia in Alzheimer's disease where the clinically observable onset is after the age of 65 years and usually in the late 70s or thereafter, with a slow progression, and usually with memory impairment as the principal feature.

Diagnostic guidelines

As for dementia, described above, with attention to the presence or absence of features differentiating the disorder from the early-onset subtype (F00.0).

F00.2 Dementia in Alzheimer's disease, atypical or mixed type

Dementias that do not fit the descriptions and guidelines for either F00.0 or F00.1 should be classified here; mixed Alzheimer's and vascular dementias are also included here.

NINCDS–ADRDA criteria for the diagnosis of Alzheimer's disease (McKhann *et al.*, 1984)[7]

Criteria for the clinical diagnosis of **probable** Alzheimer's disease include all of the following:

1. Dementia established by clinical examination and documented by the Mini-Mental Test; Blessed Dementia Scale, or some similar examination, and confirmed by neuropsychological tests.

2. Deficits in two or more areas of cognition.
3. Progressive worsening of memory and other cognitive functions.
4. No disturbance of consciousness.
5. Onset between ages 40 and 90, most often after age 65.
6. Absence of systemic disorders or other brain diseases that in and of themselves could account for the progressive deficits in memory and cognition.

The diagnosis is supported by progressive deterioration of specific cognitive functions such as language (aphasia), motor skills (apraxia), and perceptions (agnosia); impaired activities of daily living and altered patterns of behaviour; family history of similar disorders, particularly if confirmed neuropathologically; and laboratory results of: normal lumbar puncture as evaluated by standard techniques, normal pattern or non-specific changes in EEG, such as increased slow-wave activity, and evidence of cerebral atrophy on CT with progression documented by serial observation.

Other clinical features consistent with the diagnosis of **probable** Alzheimer's disease, after exclusion of causes of dementia other than Alzheimer's disease, include: plateaus in the course of progression of the illness; associated symptoms of depression, insomnia, incontinence, delusions, illusions, hallucinations, catastrophic verbal, emotional or physical outbursts, sexual disorders, and weight loss; other neurologic abnormalities in some patients, especially with more advanced disease and including motor signs such as increased muscle tone, myoclonus, or gait disorder; seizures in advanced disease; and CT normal for age.

Features that make the diagnosis of **probable** Alzheimer's disease uncertain or unlikely include: sudden, apoplectic onset; focal neurologic findings such as hemiparesis, sensory loss, visual field deficits, and incoordination early in the course of the illness; and seizures or gait disturbances at the onset or very early in the course of the illness.

Appendix 2
Research protocol

Full title of research question

To assess the clinical effectiveness and cost-effectiveness of donepezil, rivastigmine, galantamine, and memantine for Alzheimer's disease.

Clarification of research question and scope

This is an update report for donepezil, rivastigmine and galantamine (first being completed 2000) and a new report for memantine.

The aim of the review is to (a) provide a review of the clinical effectiveness and cost-effectiveness of the symptomatic treatments of donepezil, rivastigmine, and galantamine for people suffering from mild to moderately severe Alzheimer's disease; and (b) to provide a review of the clinical effectiveness and cost-effectiveness of memantine for the symptomatic treatment of moderately-severe to severe Alzheimer's disease.

The review will include the above-mentioned drugs for the treatment of Alzheimer's disease in line with their market approval for disease severity.

Evidence will focus on randomised controlled trials (RCTs) comparing the interventions with placebo, non-drug comparators, or comparisons between the interventions.

The review will be from an NHS and personal social services (PSS) perspective (costs and benefits). Baseline analysis will be limited to an NHS and PSS perspective, but where the evidence suggests there might be important costs falling on carers or other non-NHS organisations, or carer benefits, these will be noted separately, and where possible separate analysis will be reported.

Report methods

The review will be undertaken as systematically as time allows following the general principles outlined in NHS CRD Report 4. The research protocol will be updated as necessary as the research programme progresses. Any changes in the protocol will be notified to NCCHTA and NICE.

Search strategy

Electronic databases that will be searched include: Cochrane Systematic Reviews Database, Cochrane Controlled Trials Register, NHS CRD (University of York) DARE and NHS EED, Medline (Ovid), PubMed, Embase, National Research Register, Current Controlled Trials, PsycInfo, Science Citation Index, Web of Science Proceedings, BIOSIS, HTA, Clinicaltrials.gov.

Searches for donepezil, rivastigmine and galantamine will be for the period from 2000 to 2004 and will be limited to English language (searches prior to 2000 were undertaken in the previous technology assessment report and will also be used to source eligible trials for this update review). Searches for memantine will be for the period from the inception of the database until July 2004 and will be limited to English language.

Bibliographies of related papers will be assessed for relevant studies. Experts will be contacted for advice and peer review, and to identify additional published and unpublished references. Manufacturer and sponsor submissions to NICE will be searched for studies that meet the inclusion criteria.

Inclusion and exclusion criteria

Interventions include the four drugs donepezil, rivastigmine, galantamine and memantine for Alzheimer's disease.

Participants include those people diagnosed with probable Alzheimer's disease [National Institute of Neurological and Communicative Disorders and Stroke and the Alzheimer Disease and Related Disorders Association (NINCDS-ADRDA) and/or DSM-III/IV criteria] that meet the criteria for treatment with donepezil, rivastigmine, galantamine, (mild to moderately severe Alzheimer's disease, usually associated with a MMSE score of 10–26) and memantine (moderately severe to severe Alzheimer's disease):

- Although the interventions included are licensed for specific conditions, it is evident that studies use different terms to describe the same condition. For example, patients with mild to moderately severe Alzheimer's disease according to recognised criteria may be described as having mild to moderate Alzheimer's disease or having an MMSE of 10–26 in a trial report. Also patients with moderately severe to severe Alzheimer's disease may be described as moderate to severe. The review will include trials using any of these terms to describe the patients' condition, assessing any apparent differences through sensitivity analysis.
- If the MMSE cited in each trial falls outside of the range suggested above, or reports another measure of severity, a pragmatic decision will be taken to use the definition reported in individual trials, and to note any differences in the review.

Trials of participants with mixed dementia types will be included when the predominant dementia is Alzheimer's disease. Trials will not be included if the predominant dementia is not Alzheimer's disease, or the predominant dementia is not specified.

Systematic reviews of RCTs and RCTs comparing the different drugs with placebo or each other or non-drug comparators will be included in the review of effectiveness. Systematic reviews will be used as a source for RCTs and as a comparator. Any studies published as abstracts or conference presentations will be assessed for inclusion if sufficient details are presented to make appropriate decisions about the methodology of the study and the results.

If searches show that there is no evidence of the long-term effects of treatments in terms of adverse events, then controlled clinical trials meeting the other inclusion criteria and having a duration of follow-up of 12 months or more may be considered for inclusion.

For the review of memantine (for moderately severe to severe Alzheimer's disease) trials that combine memantine with either donepezil, galantamine, or rivastigmine will be included. Trials that provide memantine following on from treatment with either donepezil, galantamine, or rivastigmine will also be included.

Outcomes will focus on those that are clinically relevant to patients with Alzheimer's dementia and their carers. Primary outcome measures will include survival and measures of global functioning, cognition, function, behaviour and mood, health related quality of life. In addition, the systematic review will report information on secondary outcomes on adverse events, ability to remain independent, likelihood of admission to residential/nursing care, carer health related quality of life, and compliance (adherence) where they are reported in the included studies. Inclusion decisions will be made on primary outcome measures.

Economic evaluations of donepezil, rivastigmine, galantamine and memantine in people with Alzheimer's disease that include a comparator (or placebo) and both the costs and consequences (outcomes) of treatment will be included. Systematic reviews of economic evaluations will also be included.

Inclusion criteria will be applied by one reviewer and checked by a second reviewer, with any disagreements resolved through discussion.

Data extraction strategy

Data will be extracted from the included studies using standard tables for the clinical and cost-effectiveness studies. Data extraction will be undertaken by one reviewer and checked by a second reviewer, with any disagreements resolved through discussion.

Quality assessment strategy

The quality of included systematic reviews will be assessed using NHS CRD (University of York) criteria. Quality assessment of RCTs will be judged in accordance with chapter II.5 of CRD Report 4 (2nd Edition). Economic evaluations will be assessed using criteria recommended by Drummond and Jefferson,[39] and/or the format recommended and applied in the CRD NHS Economic Evaluation Database (CRD Report 6). Quality criteria will be applied by one reviewer and checked by a second reviewer, with any disagreements resolved through discussion.

Methods of analysis/synthesis

Clinical effectiveness will be synthesised through a narrative review with tabulation of results of included studies. Where evidence is available, the review will undertake subgroup analyses by disease severity. Data will be combined statistically if of sufficient quantity, quality and if sufficiently similar by meta-analysis using Review Manager software.

Methods for estimating qualify of life, costs and cost-effectiveness and/or cost/QALY

Published cost-effectiveness studies will be reviewed in detail, comprising a narrative review with a tabulation of results where appropriate. Cost-effectiveness studies will be identified as part of the search strategy documented above.

Where appropriate, an economic model will be devised by adapting an existing cost-effectiveness model or constructing a new one using the best available evidence to determine cost-effectiveness in a UK setting; extrapolating from shorter-term clinical data (e.g. 6-month trial data) to longer term final outcomes (i.e. modelling disease progression over time). An exploratory review of the literature indicates that product specific modelling methods have been reported to date (see Chapter 2).

Data on resource use and costs will be from the published literature, NHS sources and industry submissions where appropriate and available. The perspective of the economic analysis will be that of the NHS and PSS. As stated above, baseline analysis will be in accordance with the perspective of the NHS and PSS, and where costs and resource use related to treatment fall outside of this perspective we will report these separately where data are available.

Effectiveness data, in terms of the outcomes described in the above section, will be extracted from published trials and used in association with cost data to populate the model to obtain measures of cost-effectiveness. If available, quality of life information will be obtained from the literature or other sources to calculate cost utility estimates in terms of cost per quality-adjusted life year (QALY). From an exploratory review of the cost-effectiveness literature we have noted the use of a variety of economic endpoints, usually using the MMSE as an indicator of disease progression.

The robustness of the results to the assumptions made in the model will be examined through sensitivity analysis and/or probabilistic sensitivity analysis.

Other considerations

It is evident that clinical trials of treatments for dementia may be affected by changes in the clinical management of patients, particularly where it focuses on the longer term. As a consequence, the systematic review will indicate any major alterations in treatment (including stopping treatment, or cross-over between groups) stated by the studies and report whether outcomes are reported on an intention to treat.

Handling the company submission(s)

Industry submissions will be checked for additional studies that meet the SHTAC inclusion criteria, for data on costs and for data on the current use of donepezil, rivastigmine, galantamine and memantine. Results of cost-effectiveness analyses from industry will be compared with the SHTAC analysis, but this will not be a line by line critique of sponsor models. Any 'commercial in confidence' data taken from the industry submissions will be clearly marked in the report submitted to the HTA programme and to NICE. In addition, any information provided by others that is deemed in confidence will be marked as academic in confidence. A separate version with any such data removed will also be submitted.

Project management

It is planned to send: a final protocol to NCCHTA on 11 March 2004; an interim progress report on 17 June 2004; a complete and near-final draft to external reviewers and NCCHTA on 14 July 2004; and the final assessment report to NCCHTA on 30 August 2004.

Competing interests

None known.

External review

The technology assessment report (TAR) will be subject to external review by at least two experts acting on behalf of the NHS HTA Programme. These referees will be chosen according to academic seniority and content expertise and will be agreed with NCCHTA. We recognise that the NICE secretariat and Appraisal Committee will undertake methodological review. In addition, an external methodological referee will be asked to review the report on behalf of the HTA Programme. Referees will review a complete and near final draft of the TAR and will understand that their role is part of external quality assurance. Referees will be required to sign a copy of the NICE Confidentiality Acknowledgement and Undertaking that we will hold on file. Comments from referees and the technical lead, together with our responses will be made available to NCCHTA in strict confidence for editorial review and approval.

Appendix 3

Search strategy

Sources of information, search terms and flow chart of study identification

Databases were searched for published studies, and recently completed and ongoing research. All searches were limited to English language only.

The following strategy was used to search Medline 1966 to February 2004, and was adapted as appropriate for the remaining databases listed below.

1. Alzheimer Disease
2. alzheimer$.ti,ab.
3. 1 or 2
4. MEMANTINE/
5. memantine.ti,ab.
6. memantin.ti,ab.
7. ebixa.ti,ab.
8. axura.ti,ab.
9. memantine.rn.
10. 4 or 5 or 6 or 7 or 9
11. GALANTAMINE/
12. galantamine.ti,ab.
13. galanthamine.ti,ab.
14. galantamin.ti,ab.
15. nivalin.ti,ab.
16. nivaline.ti,ab.
17. lycoremin.ti,ab.
18. lycoremine.ti,ab.
19. reminyl.ti,ab.
20. donepezil.ti,ab.
21. aricept.ti,ab.
22. donepezil.rn.
23. rivastigmine.ti,ab.
24. rivastigmin.ti,ab.
25. exelon.ti,ab.
26. prometax.ti,ab.
27. galantamine.rn.
28. rivastigmine.rn.
29. 11 or 12 or 13 or 14 or 15 or 16 or 19 or 20 or 21 or 22 or 23 or 24 or 25 or 27 or 28
30. 10 or 29
31. 30 and 3
32. limit 31 to (english language and yr=2004)
33. from 32 keep 1-52

Additional searching

All references of articles for which full papers were retrieved were checked to ensure that no eligible studies had been missed.

Industry submissions to NICE were examined for any further studies that met the inclusion criteria (see Appendix 12).

Additional searches for cost of treatment/burden of disease, epidemiology, economic and disease modelling and quality of life were also undertaken.

Additional databases searched

	Date or issue of databases searched	
	Clinical effectiveness	**Cost effectiveness and QoL**
Cochrane Library	Issue 1, 2004	Issue 1, 2004
Embase	1980 – February 2004	1980 – February 2004
PsychInfo	1985 – February 2004	
Science Citation Index	1981 – February 2004	1981 – February 2004
ISI Web of Science Proceedings	1990 – February 2004	1990 – February 2004
BIOSIS	2000 – February 2004	2000 – February 2004
DARE	1995 – February 2004	1995 – February 2004
HTA Database	1998 – February 2004	1998 – February 2004
National Research Register	2000 – February 2004	2000 – February 2004
Current Controlled Trials	February 2004	
Clinicaltrials.gov	February 2004	
NHS EED		1995 – February 2004
EconLit		1969 – February 2004

See *Figure 24* for the flowchart of identification of studies.

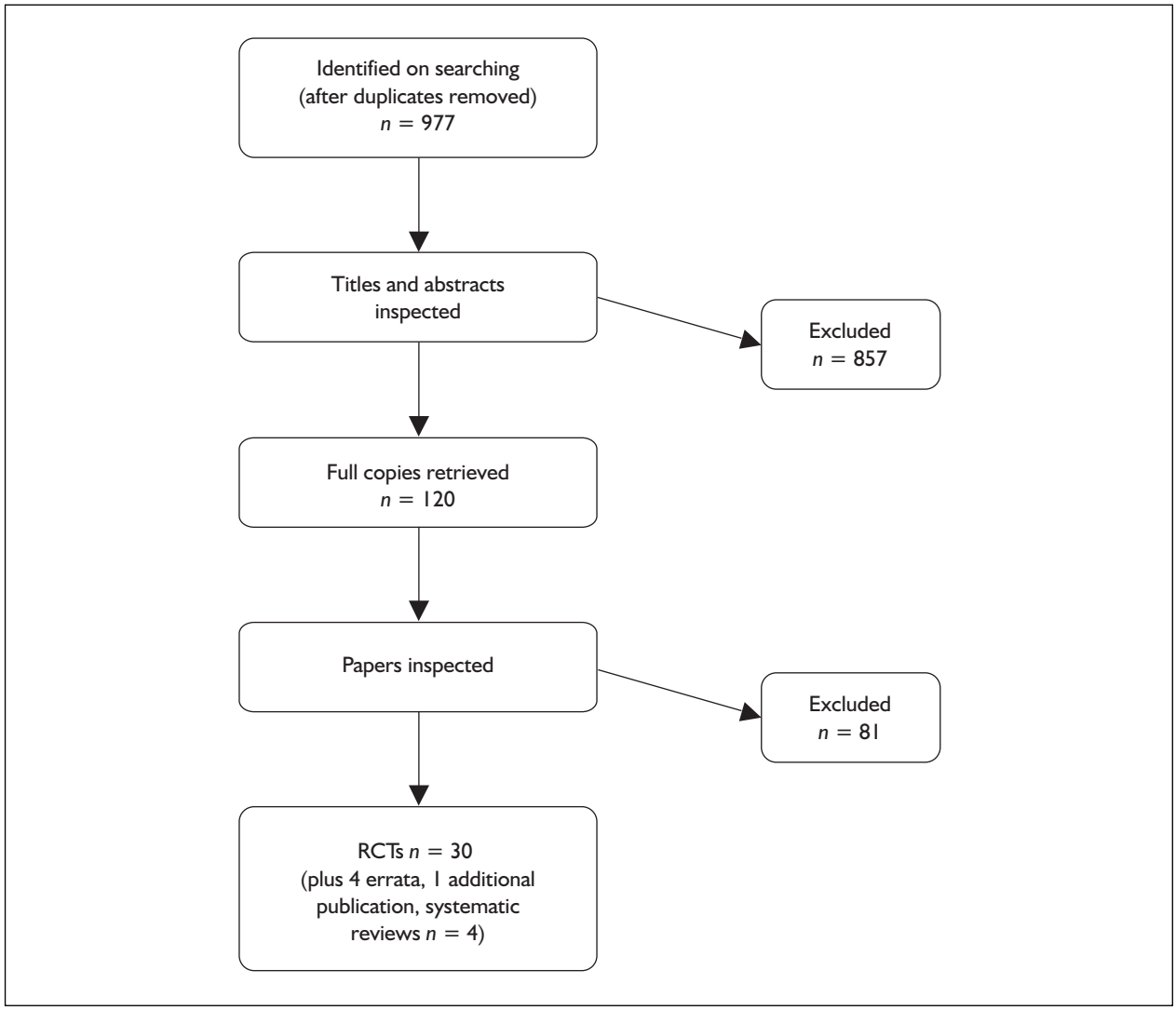

FIGURE 24 *Flowchart of identification of studies for inclusion in the systematic review of clinical effectiveness*

Appendix 4

Excluded studies

1. Double-blind trial will compare two anti-Alzheimer's drugs. *Journal of Dementia Care* 2001;**9**(5).
2. Allegri RF, Mangone CA, Duret F, Arizaga RL, Adamson J, Drake M, *et al*. Efficacy and safety of donepezil in Argentina. A 12 week, open label trial. *Revista Neurologica Argentina* 2002;**27**(1):17–23.
3. Amatniek J, Ancoli-Israel S, Lindsey L. Methodology, demographics, and preliminary safety and efficacy data from a pilot head-to-head study of the effects of galantamine (Reminyl®) and Aricept®) on sleep and attention. Poster presented at the 17th Annual Meeting of the American Association for Geriatric Psychiatry (AAGP), Baltimore, Maryland, February 21–24, 2004.
4. Anand R, Messina J, Hartman R. Dose–response effect of rivastigmine in the treatment of Alzheimer's disease. *International Journal of Geriatric Psychopharmacology* 2000;**2**(2):68–72.
5. Anand R, Farlow M, Hartman R, Graham S. Analysis of outcome in patient dropouts originally treated with rivastigmine versus placebo in a 26-week, Alzheimer's disease trial. *Neurology* 2001;**56**(8 Supplement 3):A339–A340.
6. Areosa SA, Sherriff F. Memantine for dementia. *The Cochrane Library*, Issue 1, 2004. Chichester: John Wiley, 2004.
7. Bentham P, Gray R, Hill R, Sellwood E, Courtney C. Twelve week response to cholinesterase inhibitors does not predict future benefit: the AD2000 trial experience. The 8th Conference on Alzheimer's Disease and Related Disorders, 337. Stockholm, 2002.
8. Bergman J, Brettholz I, Shneidman M, Lerner V. Donepezil as add-on treatment of psychotic symptoms in patients with dementia of the Alzheimer's type. *Clinical Neuropharmacology* 2003; **26**(2):88–92.
9. Birks JS, Harvey R. Donepezil for dementia due to Alzheimer's disease. *The Cochrane Library*, Issue 1, 2004. Chichester: John Wiley, 2004.
10. Bullock R, Erkinjuntti T, Lilienfeld S. Management of patients with Alzheimer's disease plus cerebrovascular disease: 12-month treatment with galantamine. *Dementia and Geriatric Cognitive Disorders* 2003;**17**(1–2):29–34.
11. Cummings JL, Schneider L, Tariot PN, Kershaw PR, Yuan W. Reduction of behavioural disturbances and caregiver distress by galantamine in patients with Alzheimer's disease. *Am J Psychiatry* 2004;**161**(3):532–8.
12. Cummings J, Anand R, Koumaras B, Hartman R. Rivastigmine provides behavioural benefits to Alzheimer's disease patients residing in a nursing home: Findings from a 26-week trial. *Neurology* 2000;**54**(7 Supp. 3):3.
13. Deleu D. A 1-year, randomized, placebo-controlled study of donepezil in patients with mild to moderate AD. *Neurology* 2002;**58**(5):835–6.
14. Doggrell S. Is memantine a breakthrough in the treatment of moderate-to-severe Alzheimer's disease? *Expert Opin Pharmacother* 2003; **4**(10):1857–60.
15. Dooley M, Lamb HM. Donepezil: a review of its use in Alzheimer's disease. *Drugs and Aging* 2000; **16**(3):199–226.
16. Erkinjuntti T, Kurz A, Gauthier S, Bullock R, Lilienfeld S, Damaraju CV. Efficacy of galantamine in probable vascular dementia and Alzheimer's disease combined with cerebrovascular disease: a randomised trial. *Lancet* 2002;**359**(9314):1283–90.
17. Farlow MR. Memantine/donepezil dual therapy is superior to placebo/donepezil therapy for treatment of moderate to severe Alzheimer's disease. *Neurology* 2003;**60**(5 Supplement 1):Abstract A412.
18. Fleischhacker WW, Buchgeher A, Schubert H. Memantine in the treatment of senile dementia of the Alzheimer type. *Progress in Neuro Psychopharmacology & Biological Psychiatry* 1986; **10**(1):87–93.
19. Froelich L. Donepezil for Alzheimer's Disease: the Donald study. A multicenter 24 weeks clinical trial in Germany. *European Neuropsychopharmacology* 2000;**10**(Supplement 3):S360–S361.
20. Froelich L, Moeller HJ. Two controlled clinical trials on a combination therapy of galantamine plus memantine in Alzheimer's disease and MCI-rationale and design. *Proceedings of the 8th International Conference on Alzheimer's Disease and Related Disorders*; July, Stockholm, Sweden, 2002.
21. Galasko DJ. Functional improvement from treatment with the NMDA antagonist memantine: Results of a 28-week, randomized, placebo-controlled study in advanced Alzheimer's disease. *European Journal of Neurology* 2003; **10**(Supplement 1):160.

22. Goldlist B, Gordon M, Naglie G. Galantamine vs. donepezil in the treatment of Alzheimer's disease. *Drugs & Aging* 2003;**20**(15):1139–40.

23. Grimley Evans J, Wilcock GK, Whitehead A, Birks J, Perdomo C. Donepezil: a meta-analysis of individual patient data from randomised controlled trials in the treatment of patients with mild to moderate alzheimer's disease. *JNS* 2001; **187**(Suppl. 1):57–8.

24. Grossberg G. Comparative study of donepezil and rivastigmine. *International Journal of Clinical Practice* 2003;**57**(1):70–1.

25. Homma A, Imai Y, Hagiguchi S, Hasegawa K, Kameyama M, Nishimura T. Late phase II clinical study of acetyl cholinesterase inhibitor E2020 in patients with Alzheimer-type dementia. *Clinical Evaluation* 1998;**26**(2):251–3.

26. Jarvis B, Figgitt DP. Memantine. *Drugs & Aging* 2003;**20**(6):465–76.

27. Kertesz A. Efficacy of galantamine in probable vascular dementia and Alzheimer's disease combined with cerebrovascular disease: a randomized trial. *Current Neurology & Neuroscience Reports* 2002;**2**(6):503–4.

28. Kertesz A. Memantine in moderate to severe Alzheimer's disease. *Current Neurology & Neuroscience Reports* 2003;**3**(6):485–6.

29. Kewitz H, Berzewski H, Rainer M, Dal-Bianco P, Friedl E, Deisenhammer J, *et al*. Galantamine, a selective non-toxic acetylcholinesterase inhibitor is significantly superior over placebo in treatment of SDAT. *Neuropsychopharmacology* 1994;**10**(3S):130S.

30. Kim JM, Shin IS, Yoon JS. Correlates of dropout, efficacy, and adverse events in treatment with acetylcholinesterase inhibitors in Korean patients with Alzheimer's disease. *International Psychogeriatrics* 2002;**14**(2):187–95.

31. Kozubski W, Kurz A. Donepezil-treated Alzheimer's disease patients with apparent initial cognitive decline demonstrate significant benefits when therapy is continued: results from a randomized, placebo-controlled trial. *European Neuropsychopharmacology* 2003;**13**(Supplement 4): S405.

32. Kumar V, Anand R, Messina J, Hartman R, Veach J. An efficacy and safety analysis of Exelon in Alzheimer's disease patients with concurrent vascular risk factors. *European Journal of Neurology* 2000;**7**(2):159–69.

33. Lanctot KL, Herrmann N, Yau KK, Khan LR, Liu BA, LouLou MM, *et al*. Efficacy and safety of cholinesterase inhibitors in Alzheimer's disease: a meta-analysis. *Canadian Medical Association Journal* 2003;**169**(6):557–64.

34. Lilienfeld S, Papadopoulos. Galantamine alleviates caregiver burden in Alzheimer's disease. *14th Annual Meeting of the American Association for Geriatric Psychiatry*; February, San Francisco, CA, USA, 2001.

35. Lilienfeld S, Parys W. Galantamine: additional benefits to patients with Alzheimer's disease. *Dementia & Geriatric Cognitive Disorders* 2000; **11**(Suppl1):19–27.

36. Marcusson J, Bullock R, Gauthier S, Kurz A, Schwalen S. Galantamine demonstrates efficacy and safety in elderly patients with Alzheimer disease. *Alzheimer Dis Assoc Disord* 2003;**17**:S86–S91.

37. Marder K. Efficacy of galantamine in probable vascular dementia and Alzheimer's disease combined with cerebrovascular disease: a randomized trial. *Current Neurology & Neuroscience Reports* 2002;**2**(5):389–90.

38. Mastey V, Wimo A, Winblad B, Haglund A, Jacobson L, Miceli R, *et al*. An economic evaluation of donepezil in mild to moderate Alzheimer's disease: results of a one year, double-blind, randomized trial. *Journal of the American Geriatrics Society* 2001;**49**(4):S131.

39. Mastey V, Wimo A, Winblad B, Haglund A, Jacobson L, Miceli R, *et al*. Donepezil reduces the time caregivers spend providing care: results of a one-year, double-blind, randomized trial in patients with mild to moderate Alzheimer's disease. *14th Annual Meeting of the American Association for Geriatric Psychiatry*, February, San Francisco, CA, USA, 2001.

40. Moebius HJ. Functional improvement from treatment with the NMDA antagonist memantine: results of a 28-week, randomized, placebo-controlled study in advanced Alzheimer's disease. *European Neuropsychopharmacology* 2003; **13**(Supplement 4):S388.

41. Mohs R, Doody R, Morris J, Ieni J, Rogers S, Perdomo C, *et al*. Donepezil preserves functional status and improves cognition in alzheimer's disease patients: results from a 1-year prospective placebo-controlled study. *Journal of the American Geriatrics Society* 2000;**48**(8):S46.

42. Mohs R, Doody R, Morris J, Ieni JR, Rogers SL, Perdomo CA, *et al*. Donepezil preserves functional status in Alzheimer's disease patients: results from a 1-year prospective placebo-controlled study. *Proceedings of the Quality Research in Dementia Conference*, November, London, 2000.

43. Mohs R, Doody R, Morris J, Ieni J, Pratt R, Rogers S. Donepezil preserves activities of daily living in Alzheimer's disease patients: results from a one-year placebo-controlled functional survival study. *Neurology* 2000;**54**(3):A415.

44. Passmore P, Wetterberg P, Adler G, Bullock R, Soininen H, Aarsland D, *et al*. First head to head study comparing the tolerability, ease of use, and efficacy of donepezil and galantamine in

Alzheimer's disease. *Proceedings of the 8th International Conference on Alzheimer's Disease and Related Disorders*, July, Stockholm, Sweden, 2002.

45. Peng D, Xu X, Hou Q. The safety and efficacy of aricept in patients with Alzheimer disease. *Chinese Journal of Neurology* 2002;**35**(1):19–21.

46. Perdomo C, Wilcock GK, Whitehead A, Evans JG, Birks J, Pratt RD. Donepezil: a meta-analysis of individual patient data from randomized controlled trials in the treatment of patients with mild to moderate Alzheimer's disease. *Journal of General Internal Medicine* 2003;**18**:197.

47. Pratt R, Mohs R, Doody R, Morris J, Rogers S, Ieni J, *et al*. Donepezil preserves functional status in Alzheimer's disease patients: results from a 1-year prospective placebo controlled functional study. *Proceedings of the 6th International Stockholm/Springfield Symposium on Advances in Alzheimer Therapy*, April, Stockholm, Sweden, 2000; p. 178.

48. Pratt RD, Perdomo CA, Surick IW, Ieni JR. Donepezil: tolerability and safety in Alzheimer's disease. *International Journal of Clinical Practice* 2002; **56**(9):710–17.

49. Rasmusen L, Yan B, Robillard A, Dunbar F. Effects of washout and dose-escalation periods on the efficacy, safety, and tolerability of galantamine in patients previously treated with donepezil: ongoing clinical trials. *Clinical Therapeutics* 2001;**23**: Suppl 30.

50. Reisberg B, Stoeffler A, Ferris SH, Schmitt F, Doody RS. A placebo-controlled study of memantine in advanced Alzheimer's disease. *2001 Annual Meeting of the American Psychiatric Association*, May, New Orleans; LA, USA, 2001.

51. Reisberg B, Mobius HJ, Stoffler A, Schmitt F, Doody R. Long term treatment with the NMDA antagonist memantine: results of a 24-week open-label extension study in moderately severe to severe Alzheimer's disease. *Proceedings of the 8th International Conference on Alzheimer's Disease and Related Disorders*, July, Stockholm, Sweden, 2002.

52. Reisberg B, Windscheif U, Ferris SH, Hingorani VN, Stoeffler-Moebius HJ. Memantine in moderately severe to severe Alzheimer's disease (AD): results of a placebo-controlled 6-month trial. *Proceedings of the World Alzheimer Congress*, July, Washington, USA, 2000.

53. Reisberg B. Results of a placebo-controlled 6-month trial with memantine in moderate to severe Alzheimer's disease (AD). *European Neuropsychopharmacology* 2000;**10**(Supplement 3): S363–4.

54. Reisberg B, Windscheif U, Ferris SH, Stoeffler A, Moebius HJ, The Memantine Study Group. Treatment of advanced alzheimer's disease with memantine, an NMDA antagonist: results of a 6-month multicenter randomized controlled trial. *39th Annual Meeting of the American College of Neuropsychopharmacology*, December, San Juan, Puerto Rico, 2000.

55. Reisberg B, Stoeffler AF-S. A placebo-controlled study of memantine in advanced Alzheimer's disease. *15th Annual Meeting of the American Psychiatric Association*, May, Philadelphia, PA, USA, 2002.

56. Rogers SL, Doody RS, Pratt RD, Ieni JR. Long-term efficacy and safety of donepezil in the treatment of Alzheimer's disease: final analysis of a US multicentre open-label study. *European Neuropsychopharmacology* 2000;**10**(3):195–203.

57. Rozzini L, Bargnani C, Bosio A, Chia F, Franzoni S, Leonardi R, *et al*. Comparison of efficacy and safety of rivastigmine and donepezil in patients with mild to moderate alzheimer disease: results from a multicentre randomised trial. *Proceedings of the 7th International Geneva/Springfield Symposium on Advances in Alzheimer Therapy*, April, Geneva, 2002; p. 240.

58. Santens P, Ventura M. Donepezil in the treatment of mild to moderate Alzheimer's disease: report of a Belgian multicenter study. *Acta Neurologica Belgica* 2003;**103**(3):159–63.

59. Shua HJ, Smith J. A head to head study of donepezil (aricept), rivastigmine (exelon) and galantamine (reminyl) for the treatment of Alzheimer's disease: safety, tolerability, clinical and caregiver impression after 4–5 months of treatment (a prospective study). *Proceedings of the 8th International Conference on Alzheimer's Disease and Related Disorders*, July, Stockholm, Sweden, 2002.

60. Shua-Haim JR, Pass MD, Smith J, Lee PP, Patel S. An investigation into the safety, tolerability, and efficacy of high dose rivastigmine treatment in Alzheimer's disease. *Journal of the American Geriatrics Society* 2004;**52**(4):223.

61. Smith J, Shua-Haim JR, Pass MD, Lee PP, Patel P. Results of a next day crossover study of donepezil to galantamine in Alzheimer's disease patients: A one-month follow up study. *Journal of the American Geriatrics Society* 2004;**52**(4):224.

62. Tai CT, Liu CK, Sung SM, Pai MC, Hsu CY. The safety and efficacy of exelon in Alzheimer's patients: a multicentre, randomized, 26-week study in Taiwan. *International Journal of Neuropsychopharmacology* 2000;**3**(Supplement 1): S356.

63. Tariot P, Perdomo CA, Whalen E, Sovel MA, Scham EM. Age is not a barrier to donepezil treatment of Alzheimer's disease in the long-term care setting. *International Psychogeriatrics* 1999; **11**:134.

64. Tariot P. Memantine/donepezil dual-therapy is superior to placebo/donepezil therapy for treatment of moderate to severe Alzheimer's disease. *European Journal of Neurology* 2003;**10**(Supplement 1):25–6.

65. Tariot P, Parys W. The efficacy and tolerability of galantamine in alzheimer's disease: a 5 month placebo controlled study with slow dose escalation. *Neurology* 2000;**54**(Suppl 3):A415.

66. Tariot PN, Cummings JL, Katz IR, Mintzer J, Perdomo CA, Schwam EM, *et al*. A randomised, double-blind, placebo-controlled study of the efficacy and safety of Donepezil in patients with Alzheimer's disease in the nursing home setting. *Journal of the American Geriatrics Society* 2001; **49**(12):1590–9.

67. Tariot R, Farlow M, Grossberg G, Gergel I, Graham S, Jin J. Memantine/donepezil dual-therapy is superior to placebo/donepezil therapy for treatment of moderate to severe Alzheimer's disease. *Journal of the American Geriatrics Society* 2003;5**1**(4):542.

68. Triau E, Johannsen P. Donepezil-treated Alzheimer's disease patients with apparent initial cognitive decline demonstrate significant benefits when therapy is continued: results from a randomized, placebo-controlled trial. *European Journal of Neurology* 2003;**10**(Supplement 1):23–4.

69. Trinh N-H, Hoblyn J, Mohanty S, Yaffe K. Efficacy of cholinesterase inhibitors in the treatment of neuropsychiatric symptoms and functional impairment in Alzheimer disease: A meta-analysis. *Journal of the American Medical Association* 2003; **289**(2):210–6.

70. Van Dyck C, Ferris S, Graham S, Stoeffler A, Olin J, Wirth Y, *et al*. Response to memantine treatment in moderate to severe Alzheimer's disease. *Journal of the American Geriatrics Society* 2004;**52**(4):405.

71. Van Gool WA, Rogers SL, Friedhoff LT, Farlow MR, Doody RS, Mohs R. Efficacy of donepezil in Alzheimer's disease: fact or artifact? *Neurology* 1999; **52**(1):218–19.

72. Wang Y, Chen Q, Zhang Z, *et al*. The treatment by using rivastigmine for patients with Alzheimer disease: Results of a multicenter, randomized, open-labeled, controlled clinical trial. *Chinese Journal of Neurology* 2001;**34**(4):210–13.

73. Wilcock G, Howe I, Coles H, Lilienfeld S, Truyen L, Zhu Y, *et al*. A long-term comparison of galantamine and donepezil in the treatment of Alzheimer's disease. *Drugs & Aging* 2003; **20**(10):777–89.

74. Wilcock G, Wilkinson D. Galanthamine hydrobromide: interim results of a group comparative, placebo controlled study of efficacy and safety in patients with a diagnosis of senile dementia of the Alzheimer type. In Iqbal K, Winblad B, Nishimura T, Takeda M, Wismewski HM, eds. *Alzheimer's Disease: Biology, Diagnosis and Therapeutics*, p. S144. Chichester: John Wiley, 1997.

75. Wilcock GK. Donepezil: a meta-analysis of individual patient data from randomized controlled trials in the treatment of patients with mild to moderate Alzheimer's disease. *Neurology* 2001; **56**(8 Supplement 3):A339.

76. Wilkinson D, Passmore P, Potocnik F, Maud C. Donepezil compared to rivastigmine in Alzheimer's disease: similar efficacy but better tolerability and physician and caregiver satisfaction in a multinational randomized trial. *14th Annual Meeting of the American Association for Geriatric Psychiatry*, February, San Francisco, CA, USA, 2001.

77. Wilkinson DG, Hock C, Farlow M, van Baelen B, Schwalen S. Galantamine provides broad benefits in patients with 'advanced moderate' Alzheimer's disease (MMSE ≤12) for up to six months. *International Journal of Clinical Practice* 2002; **56**(7):509–14.

78. Wilkinson DG. Comparison of the tolerability, ease of use, and efficacy of donepezil and rivastigmine in Alzheimer's disease patients: a 12-week multinational study. *Neurology* 2001; **56**(8 Supplement 3):A456–7.

79. Williams BR, Nazarians A, Gill MA. A review of rivastigmine: a reversible cholinesterase inhibitor. *Clinical Therapeutics* 2003;**25**(6):1634–53.

80. Winblad B, Poritis N, Mobius HJ. Clinical improvement in a placebo-controlled trial with memantine in care-dependent patients with severe dementia (M-BEST). *Alzheimer's Disease and Related Disorders – Etiology, Pathogenesis and Therapeutics* 1999;633–40.

The M-Best study by Winblad and Poritis (1999) was excluded from this review on the grounds that the patient group was not exclusive to those with AD. The HIS (Hachinski Iscemic Scale) was used in this study to separate patient subgroups. Although this eliminated those patients without VAD, there was no stratification to AD.

81. Winblad B. Donepezil is a well-tolerated long-term treatment for patients with mild to moderate Alzheimer's disease: results from a 1-year placebo-controlled study and 2-year follow-up study. *European Journal of Neurology* 2003; **10**(Supplement 1):158.

82. Winblad B. Efficacy and tolerability of memantine in nursing home patients with moderate-to-severe dementia of the Alzheimer's type. *European Journal of Neurology* 2003;**10**(Supplement 1):159.

83. Winblad B. Efficacy and tolerability of memantine in nursing home patients with moderate to severe dementia of the Alzheimer's type. *European Neuropsychopharmacology* 2003; **13**(Supplement 4):S394.

84. Winblad B, R, Waldemar G. Long-term efficacy of donepezil in patients with mild to moderate Alzheimer's disease: Results from a one-year placebo-controlled study and two-year follow-up study. *European Neuropsychopharmacology* 2003; **13**(Supplement 4):S404.

85. 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). *International Journal of Geriatric Psychiatry* 1999;**14**(2): 135–46.

Appendix 5

Quality assessment tool for RCTs

Quality criteria for assessment of experimental studies: CRD Report 4

1. Was the assignment to the treatment groups really random?	
2. Was the treatment allocation concealed?	
3. Were the groups similar at baseline in terms of prognostic factors?	
4. Were the eligibility criteria specified?	
5. Were outcome assessors blinded to the treatment allocation?	
6. Was the care provider blinded?	
7. Was the patient blinded?	
8. Were the point estimates and measure of variability presented for the primary outcome measure?	
9. Did the analyses include an intention-to-treat analysis?	
10. Were withdrawals and dropouts completely described?	

Some instructions for using a checklist for RCTs

Quality item	Coding	Explanation
1. Was the assignment to the treatment groups really random?		
Random sequence generation	**Adequate** **Partial** **Inadequate** **Unknown**	**Adequate:** random numbers table or computer and central office or coded packages **Partial:** (sealed) envelopes without further description or serially numbered opaque, sealed envelopes **Inadequate:** alternation, case record number, birth date, or similar procedures **Unknown:** just the term 'randomised' or 'randomly allocated', etc.

continued

Quality item	Coding	Explanation
2. Was the treatment allocation concealed?		
Concealment of randomisation The person(s) who decide on eligibility should not be able to know or be able to predict with reasonable accuracy to which treatment group a patient will be allocated. In trials that use good placebos this should normally be the case, however different modes or timing of drug administration in combination with the use of small block sizes of known size may present opportunities for clinicians who are also involved in the inclusion procedure to make accurate guesses and selectively exclude eligible patients in the light of their most likely treatment allocation; in centres with very low inclusion frequencies combined with very brief follow-up times this my also present a potential problem because the outcome of the previous patient may serve as a predictor of the next likely allocation.	**Adequate** **Inadequate** **Unknown**	**Adequate:** when a paper convinces you that allocation cannot be predicted [separate persons, placebo really indistinguishable, clever use of block sizes (large or variable)]. Adequate approaches might include centralised or pharmacy-controlled randomisation, serially numbered identical containers, on-site computer based system with a randomisation sequence that is not readable until allocation, and other approaches with robust methods to prevent foreknowledge of the allocation sequence to clinicians and patients **Inadequate:** this option is often difficult. You have to visualise the procedure and think how people might be able to circumvent it. Inadequate approaches might include use of alternation, case record numbers, birth dates or week days, open random numbers lists, serially numbered envelopes (even sealed opaque envelopes can be subject to manipulation) and any other measures that cannot prevent foreknowledge of group allocation **Unknown:** no details in text. Disagreements or lack of clarity should be discussed in the review team
3. Were the groups similar at baseline regarding the prognostic factors?		
Baseline characteristics Main aim is to enable the reviewer to see which patients were actually recruited. It enables one to get a rough idea on prognostic comparability. A real check on comparability requires multivariable stratification (seldom shown).	**Reported** **Unknown**	Consult the list of prognostic factors or baseline characteristics (not included in this appendix). Reviewer decides
4. Were the eligibility criteria specified?	**Adequate** **Partial** **Inadequate** **Unknown**	
Prestratification Consult the list of prognostic factors or baseline characteristics (not included in this appendix).	**Adequate** **Partial** **Inadequate** **Unknown**	**Single centre study** **Adequate:** prestratification on at least one factor from the list or no prestratification if the number of patients exceeds a prespecified number **Partial:** leave judgement to reviewer **Inadequate:** stratification on a factor(s) not on our list or no stratification whereas the number of patients is less than the prespecified number **Unknown:** no details in text and no way to deduce the procedure from the tables. **Multicentre study** **Adequate:** must prestratify on centre. Within each centre the criteria for single centre studies also apply **Partial:** impossible option **Inadequate:** no prestratification on centre or violating the criteria for single centre studies (see above) **Unknown:** no details in text and no way to deduce the procedure from the tables

continued

Quality item	Coding	Explanation
5. Were outcome assessors blinded to the treatment allocation?		
Blinding of assessors The assessor may be the patient (self-report), the clinician (clinical scale, blood pressure, etc.) or, ideally, a third person or a panel. Very important in judgement of cause of death but unimportant in judgement of death.	**Adequate** **Inadequate** **Unknown**	**Adequate:** independent person or panel or (self) assessments in watertight double-blind conditions **Inadequate:** clinician is assessor in trial on drugs with clear side effects or a different influence on lab results, ECGs, etc. **Unknown:** no statements on procedures and not deducible
6. Was the care provider blinded?		
Blinding of caregivers Look out for good placebos (see, hear, taste, feel, smell), tricky unmasking side effects accounting for the subjectivity of the outcome measurements and the accessibility of co-interventions by the caregivers.	**Adequate** **Partial** **Inadequate** **Unknown**	**Adequate:** placebo described as 'indistinguishable' and procedures watertight (use your imagination with the 'cheat' in mind; e.g. statement that sensitive/ unmasking lab results were kept separate from ward personnel) **Partial:** just 'double-blind' in text and no further description of procedures or nature of the placebo **Inadequate:** wrong placebo (e.g. fructose in trial on ascorbic acid) **Unknown:** no details in text
Co-interventions Register when they may have an impact on any of the outcome phenomena. Consult the list of cointerventions (not included in this appendix).	**Adequate** **Partial** **Inadequate** **Unknown**	**Adequate:** percentages of all relevant interventions in all groups **Partial:** one or more interventions omitted or omission of percentages in each group **Inadequate:** not deducible **Unknown:** no statements
7. Was the patient blinded?		
Blinding of patients This item is hard to define. Just the statement 'double-blind' in the paper is really insufficient if the procedure to accomplish this is not described or reasonably deducible by the reviewer. Good placebos (see, hear, taste, feel, smell), tricky unmasking side effects accounting for the subjectivity of the outcome measurements and the accessibility of co-interventions by the patient are required.	**Adequate** **Partial** **Inadequate** **Unknown**	**Adequate:** placebo described as 'indistinguishable' and procedures watertight **Partial:** just 'double-blind' in text and no further description of procedures or nature of the placebo **Inadequate:** wrong placebo **Unknown:** no details in text
Compliance Dosing errors and timing errors.	**Adequate** **Partial** **Inadequate** **Unknown**	**Adequate:** Medication Event Monitoring System (MEMS or eDEM) **Partial:** blood samples, urine samples (use of indicator substances) **Inadequate:** pill count or self report **Unknown:** not mentioned
Check on blinding Questionnaire for patients, caregivers, assessors and analysis of the results; the (early) timing is critical because the treatment effect may be the cause of unblinding, in which case it may be used as an outcome measure.	**Reported** **Unknown**	Reviewer decides

continued

Quality item	Coding	Explanation
8. Were the point estimates and measure of variability presented for the primary outcome measure?		
Results for the primary outcome measure	**Adequate** **Partial** **Inadequate** **Unknown**	**Adequate:** mean outcome in each group together with mean difference and its standard error (SE) or standard deviation (SD) or any CI around it or the possibility to calculate those from the paper. Survival curve with log rank test and patient numbers at later time points **Partial:** partially reported **Inadequate:** no SE or SD, or SD without N (SE = SD/N) **Unknown:** very unlikely
9. Did the analysis include an intention to treat analysis?		
Intention-to-treat analysis (ITT) Early dropout can make this very difficult. Strictest requirement is sensitivity analysis including early dropouts.	**Adequate** **Inadequate**	Reviewers should not just look for the term ITT but assure themselves that the calculations were according to the ITT principle
Dealing with missing values The percentage missing values on potential confounders and outcome measurements (seldom given) is a rough estimate of a trial's quality. One can carry them forward, perform sensitivity analysis assuming the worst and best case scenarios, use statistical imputation techniques, etc. Note that the default option (deletion) assumes that the value is randomly missing, which seems seldom justified.	**Adequate** **Partial** **Inadequate** **Unknown**	**Adequate:** Percentage of missing values and distribution over the groups and procedure of handling this stated **Partial:** some statement on numbers or percentages **Inadequate:** wrong procedure (a matter of great debate) **Unknown:** no mentioning at all of missing and not deducible from tables
Loss to follow-up This item examines both numbers and reasons; typically an item that needs checking in the methods section and the marginal totals in the tables. Note that it may differ for different outcome phenomena or time points. Some reasons may be reasons given by the patient when asked and may not be the true reason. There is no satisfactory solution for this.	**Adequate** **Partial** **Inadequate** **Unknown**	**Adequate:** number randomised must be stated. Number(s) lost to follow-up (dropped out) stated or deducible (from tables) for each group and reasons summarised for each group. **Partial:** numbers, but not the reasons (or vice versa) **Inadequate:** numbers randomised not stated or not specified for each group **Unknown:** no details in text

Appendix 6

Severity rating scales and outcome measures

Global outcome measures

Type	Construct measure and scoring	Critical appraisal
Clinical Dementia Rating (CDR) and Clinical Dementia Rating Sum of Boxes (CDR-SB)	Cognitive impairment in memory, orientation, judgement/problem-solving, community affairs, home/hobbies, and personal care 0 = none, 0.5 = questionable, 1 = mild, 2 = moderate, 3 = severe CDR-SB is a modified form which sums the ratings in the six performance categories to give a global dementia ranking	Provides physicians with a global rating that encompasses a broad range of patient characteristics and can be used by neurologists, psychiatrists and psychologists and focuses on cognition, not on items that may be related to other medical, emotional or social conditions. Good inter-rater reliability and fair to good concurrent validity. Although no work has been done on test–retest reliability, nothing so far suggests that researchers should avoid this scale when trying to stage AD. The CDR can be used as an eligibility criterion for trial participation or as an outcome measure
Global Deterioration Scale (GDS)	Progressive stages of cognitive impairment 1 (no cognitive decline) – 7 (very severe cognitive decline)	Most frequently used but ratings can misstate a patient's severity. Problems might arise when the GDS is used as an inclusion criterion for participation in an RCT. The ability to enrol desired patients could be threatened if the GDS misidentifies the stages of dementia. The GDS should not be used to stage dementia in Alzheimer's disease drug trials
Clinical Global Impression of Change scale (CGIC) and the global improvement index with interviewing of patients. Clinician Interview-Based Impression of Change (CIBIC) and with caregiver input (CIBIC-M or –Plus)	Overall improvement in patient health status assessed by clinician (–with caregiver) 1 (very much improved) – 7 (very much worse) A number of different variations are available Scale is non-parametric and of a non-interval nature	Fair to good test–retest and inter-rater reliability and concurrent validity. Results may arise from fact that groups providing global assessments do not base their ratings on the same domains. Physicians take clinical psychopathology as the basis of determining global improvement, nurses believe the amount of work needed to care for patients was important. This instrument also includes a caregiver opinion, results may differ depending on whether the rater first interviews the patient or caregiver. The number of different variations may have reduced the validity
Gottfries-Bråne-Steen (GBS)	Motor function, intellectual function, emotional function and symptoms common to demented patients. 0 (normal function or absence of symptoms) to 6 (maximal disturbance or presence of symptoms)	Psychometric properties range from fair to good. Scale is useful mean of quantifying dementia in drug trials. GBS should not be used as a diagnostic tool
Mental Function Impairment Scale (MENFIS)	A modification of the GBS prepared by the study authors for a previous study. Scores range from 0 to 78, with a higher score indicating a greater degree of deficit	Unable to source data on reliability and validity
[Commercial/ academic confidential information removed]	**[Commercial/academic confidential information removed]**	**[Commercial/academic confidential information removed]**

Cognitive outcome measurements scales

Type	Construct measure and scoring	Critical appraisal
Alzheimer's Disease Assessment Scale-cognitive (ADAS-cog)	Orientation, memory, language and praxis 0–70, with higher scores indicating greater impairment	Limited in its ability to detect change at one end or the other of the severity continuum. For many subtests, detection of improvement appears only possible for a restricted range of severity levels. Limitations should be considered when used as a drug efficacy measure. The rate of decline of AD using ADAS-cog suggests that the decline is non-linear and not a constant but is dependent on the stage of the disease. Content and ecological validity are lacking
Benton Visual Retention Test (BVRT)	Assesses visual perception, visual memory and visuoconstructive abilities. The test has three alternate forms, each consisting of ten designs. In addition, there are four possible modes of administration. Scoring is based on an assessment of the number and types of errors made compared with the expected scores found in the norm tables. The wider the discrepancy in favour of the expected score, the more probable it is that the participant has suffered neurological impairment	The interscorer agreement for total error score is high and for major categories of errors reliability is moderate to high. A correlation of 0.42 was found between the Benton and the Digit Span Wechsler Adult Intelligence Scale subtest. This low correlation indicates discriminate validity since the Benton was created to supplement the Digit Span test. Educational level may influence a participant's score on the test. Participants with higher educational levels tend to use a more exhaustive exploration strategy during the recognition phase of the test, allowing them to perform better than participants with lower educational levels. The executive working memory component is more efficient in participants with higher educational levels
Computerised Memory Battery (CMBT)	A computerised version of the Memory Assessment Clinical Battery (MAC) designed to simulate critical cognitive tasks: Name-Face Association (delayed recall and total acquisition); First and Last Names (total acquisition); Facial Recognition (first miss and total correct); Telephone Number Recall (7-digit and 10-digit number correct); House and Object Placement Task (total acquisition and first trial)	The MAC-Q questionnaire demonstrates internal consistency and test-retest reliability
Digit Symbol Substitution Subtest (DSST) of the Wechsler Adult Intelligence Scale-Revised	Participants fill in a grid of 100 blank squares, each paired with a randomly assigned number from 1 to 9, using a key that pairs each number with a different symbol. The score is the number of correct answers after 90 seconds	Performance on this test is affected by many different components, so the test lacks specificity. Participants with impaired vision or visuomotor coordination, pronounced motor slowing or low education levels are at a disadvantage
Fuld Object-Memory Evaluation (FOME)	Ten item assessment with ten common objects in a bag are presented "to determine whether the patient can identify objects by touch" (stereognosis). The test was developed while testing large samples of aged adults, nursing home residents and community active people, for whom norms are provided	Unable to source data on reliability and validity

continued

Type	Construct measure and scoring	Critical appraisal
Mini-Mental State Examination (MMSE)	11 questions on orientation, memory, concentration, language and praxis. Scale ranges from 0 to 30. Higher score indicates less impairment. There is no range of scores that can be rigidly and universally applied to indicate dementia severity, i.e. as a marker of mild, moderate and severe dementia. In clinical trials often a score of 21–26 is associated with mild AD, moderate AD is associated with an MMSE of 10–20 and severe AD is usually associated with an MMSE of less than 10. This may be less suitable within routine daily practice	Good reliability and validity for its original purpose of screening for dementia, short screening scales are not designed to measure more subtle aspects of cognition. Short scales such as the MMSE may indicate little or no change over time in subjects who would otherwise be shown to have declined substantially if another scale had been used to measure change in status. Not an ideal outcome measure for AD drug trials, especially if the expected benefits are not large. It has dependence on intact language ability and there are no available validated versions in languages suitable for use with ethnic minorities. It cannot be used effectively in people with low IQs or learning disabilities
Severe Impairment Battery (SIB)	A measure of cognition that was developed to assess a range of cognitive functioning in individuals who are too impaired to complete standard neuropsychological tests and takes into account specific behavioural and cognitive deficits associated with severe dementia. It is composed of 40 simple one-step commands which are scored on a three-point scale and are presented in conjunction with gestural cues. The SIB also allows for non-verbal and partially correct responses. The six major subscales are attention, orientation, language, memory, visuo-spatial ability, and construction. Overall scores range from 0 to 1000 with positive scores indicating clinical improvement	The SIB has been shown to be psychometrically reliable and clinical norms are available. No further details of reliability and validity have been sourced
Trail Making Test (TMT)	Assesses speed of visual search, attention, mental flexibility and motor function. The test has two parts: (A) drawing a line linking numbers in sequence and (B) drawing a line linking letters in sequence. The reviewer calls any mistakes to the attention of the participant, and these must be corrected before progressing. The score is the time taken to successfully complete a test	Reliability is reported to be higher for part A than for part B, which requires more information-processing ability and is more sensitive to brain damage. Reliability is restricted due to the use of time scores rather than both error counts and time scores, since error correction may take longer in some participants than others. Scores are strongly affected by the participant's education level
Wechsler logical memory test	This test is one of 13 subtests of the Wechsler Memory Scale-Revised. The first subtest is for screening purposes, and the other 12 are grouped into five separate memory areas. The test manual provides guidelines for scoring and weighting, and provides norms for individuals aged 16–74 with information about significant differences between any two scores	Test–retest reliability and concurrent validity with a verbal learning test are adequate for the whole WMS-R test. Level of education affects a participant's score. Normative data for those aged 75 and over is lacking. The score is more heavily influenced by verbal memory performance than by other memory components

Functional and quality of life outcome measurement scales

Type	Construct measure and scoring	Critical appraisal
Alzheimer's Disease Cooperative Study-Activities of Daily Living (ADCS/ADL)	This rating scale is a 23-item assessment of ADLs that is scored from 0 (greatest impairment) to 78. It evaluates activities of daily living	The ADCS/ADL is a structured questionnaire originally created to assess functional capacity over a broad range of severity of dementia. The ADAS/ADL19 is a subset of the original inventory and focuses on items appropriate for the assessment of later stages of dementia. The sensitivity and reliability of this modification has been established
Alzheimer's Disease Functional Assessment and Change Scale (ADFACS)	Scale consists of 10 items for instrumental ADL: ability to use the telephone, performing household tasks, using household appliances, handling money, shopping, preparing food, ability to get around both inside and outside the home, pursuing hobbies and leisure activities, handling personal mail, grasping situations or explanations. Scale has a range of 0–54 where lower scores correspond to better function. Test takes approximately 20 minutes to complete	Full assessment of psychometric properties not yet published. Has face validity for those with mild-moderate AD. The ADL items chosen for this scale have been demonstrated to be sensitive to change over 12 months, correlate well with MMSE scores, and have good test-retest reliability (although several questions have been modified in the scale)
Behavioural Rating Scale for Geriatric Patients (BGP)	Consists of 35 items (scored 0, 1, or 2) assessing observable aspects of cognition, function and behaviour. A high score indicates worse function	Unable to source data on reliability and validity
Bristol Activities of Daily Living scale (BADL)	Caregiver assessment of 20 ADLs. Categories included are food, eating, drinks, drinking, dressing, hygiene, teeth, bath, toilet, transferring, mobility, orientation to time and space, communication, telephone, housework/gardening, shopping, finances, hobbies and transport. Scores range from 0 to 60 with higher scores indicating better function	Designed specifically for use with patients with dementia. Face validity was measured by asking carers whether items were important, and construct validity was confirmed by principal components analysis. Concurrent validity was assessed by observed performance, the test has good content validity, and there is good test-retest reliability. The test is shown to correlate well with performance ADLs and tests of cognitive function
Caregiver-rated Modified Crichton Scale (CMCS)	A modified Crichton Geriatric Rating Scale (CGRS). This a seven-item scale using a Likert-type scoring method. Questions include comprehension to time and place, carrying out conversation, cooperation, restlessness, dressing, social activities and leisure. Negative change relates to clinical improvement	Reliability demonstrated. Unable to source data on validity
Disability Assessment for Dementia (DAD)	This rating scale is a 46-item structured interview or questionnaire for the caregiver that is scored from 0 to 100 (least impairment). It evaluates ADLs and takes approximately 20 minutes to complete. It is based on a recognised conceptual definition of disability from the WHO	The DAD scale demonstrates a high degree of internal consistency and excellent interrater and test-retest reliability. Full details of concurrent and construct validity not yet published
Functional Assessment Staging Scale (FAST)	Assesses the magnitude of progressive functional deterioration in patients with dementia by identifying characteristic progressive disabilities. Seven major stages range from normal (stage 1) to severe dementia (stage 7)	FAST has been shown to be a reliable and valid assessment technique for evaluating functional deterioration in AD patients throughout the entire course of the illness. Because the elements of functional capacity incorporated in FAST are relatively universal and readily ascertainable, as well as characteristic of the course of AD, FAST can serve as a strong diagnostic and differential diagnostic aid for clinicians

continued

Type	Construct measure and scoring	Critical appraisal
General Health Questionnaire (GHQ-30)	The GHQ is a self-report psychiatric screening test, and items include questions on: depression and unhappiness, anxiety and felt psychological disturbance, social impairment and hypochondriasis. Participants rate themselves on a four-point severity scale, according to how they have recently experienced each GHQ item: better than usual, same as usual, worse than usual, or much worse than usual. Normally each item is scored either 0 or 1, depending on which severity choice is selected. Individual items are summed to give the total score	GHQ-30 is based on Medical Outcomes Study Short Form-36, which is extensively validated
Instrumental Activities of Daily Living (IADL)	For women, the set of behaviours assessed includes telephoning, shopping, food preparation, housekeeping, laundering, use of transport, use of medicine and ability to handle money. For men, the areas of food preparation, housekeeping and laundering are excluded. Each of the behavioural areas is given a score of 0 or 1, leading to an overall score that ranges from 0 to 8 for women and from 0 to 5 for men	The IADL is a very frequently used and often cited instrument for assessing the instrumental competence of elderly patients. The scale is well anchored from a theoretical point of view and the behaviours that are included are likely to be affected in the first stages of dementia
The Interview for Deterioration in Daily Living in Dementia (IDDD)	The IDDD measures functional disability in self-care (16 items such as washing, dressing and eating) and complex activities (17 items such as shopping, writing, and answering the telephone) Severity of impairment is rated on a 7-point scale, where 1–2 = no or slight impairment, 3–4 = mild impairment, 5–6 = moderate impairment, 7 = severe impairment, giving a total range score of 22–231	This scale appears to be appropriate to assess community-living patients with mild and moderate levels of dementia. It assesses a substantial proportion of complex activities likely to be affected during the first stages of the AD. The number of non-redundant items in the scale is viewed positively since it may increase the sensitivity of the tool. Empirical information on the testing of the IDDD and its measurement properties is seriously lacking
Physical Self-Maintenance Scale (PSMS)	Measured through competence of 6 behaviours: toileting, feeding, dressing, grooming, locomotion and bathing. It can be completed by untrained staff based on information from subjects, caregivers, friends etc. Each behavioural area is given a score of 1 or 0, with over score ranging from 0 to 6. Using Guttman scaling, each scale point has 5 descriptive scale points	Brief assessment of activities of daily living. Theoretically well grounded, it has been proven useful for evaluation of institutionalised elderly but has a ceiling effect for those living in the community. Testing of psychometric properties is incomplete
The Progressive Deterioration Scale (PDS)	PDS examines activities of daily living and instrumental activities of daily living. Examples are: extent to which a patient can leave the immediate neighbourhood, use of familiar household implements, involvement in family finances, budgeting. Each question is scored by measuring the distance along the line on a scale from 0 to 100, with higher scores reflecting better functionality. A composite score is derived from averaging across the items for a maximal score of 100. The scale is sometimes classified as a measure of quality of life	This scale has been shown to be sensitive to three severity stages of dementia although there has been some debate whether the content is adequate to assess those with moderately severe AD. The scale was systematically developed and tested on a fairly large sample of AD patients (although the mean age of the final test group was only 69.5 years). Test–retest reliability was determined in 123 patients, giving stage correlations (rs) of 0.889 for early AD (14 participants), 0.775 for 44 middle stage participants and 0.775 for 65 late stage participants. A moderate degree of correlation has been demonstrated between PDS and ADAS-cog scores (rp = –0.57 to –0.64)

continued

Type	Construct measure and scoring	Critical appraisal
		There is considerable reduplication within the scale – 4 questions relate to handling finances but there are no items pertaining to basic activities such as washing, dressing and toileting. The scale is therefore not thought to have adequate content to assess people with moderately severe AD as it does not assess the wide range of daily living skills affected at different stages of the disease. There are high levels of between and within patient variability (in the order of 12 points) which may make it less suited to detect differences over short time periods
QoL (patient and caregiver scales)	This assessment was a 7-item patient-rated scale evaluating the patients' perceptions of their well-being in terms of relationships, eating and sleeping, and social and leisure activities. The test is conducted by interview. Scored on an analogue scale between 0 (worst quality) to 50 (best quality)	This instrument has not been validated in patients with Alzheimer's disease but was selected because no QoL instrument has been validated in this population
Unified Activities of Daily Living Form (Unified ADL)	All self-care and mobility variables commonly used to assess patient's functional status. A 20-item scale was produced. The need for assistance is scored for every item, on a 10-point scale	The psychometric properties of this scale, resulting from the combination of existing evaluations, have not been published

Behaviour and mood outcome measurements scales

Type	Construct measure and scoring	Critical appraisal
NOSGER – Nurses' Observation Scale for Geriatric Patients	Contains 30 items of behaviour, each rated on a 5-point scale according to frequency of occurrence. Item scores are summarised into 6 dimension scores (memory, instrumental activities of daily life, self-care, mood, social behaviour and disturbing behaviour)	This scale has been validated, and has high inter-rater and test–retest reliability. The test correlates well with clinician's global rating of change
Neuropsychiatric Inventory (NPI)	Currently evaluates 12 items: delusions, hallucinations, dysphoria, anxiety, agitation, euphoria, apathy, irritability, disinhibition, aberrant motor behaviour, night-time behaviour and changes in appetite/eating behaviour. Psychometric properties were established on first 10 items. Total score for each domain is calculated by multiplying frequency rating by severity rating, adding domain scores to get a total score. Higher scores represent more problems. Maximum scores is 12 per domain, with either 10 or 12 domains assessed	Content validity has been established, reliability and validity are satisfactory. Limitations included: poor description of appraisal period for behavioural symptoms; no justification for scoring system; and, inter-rater reliability was poorly deserved

Appendix 7

Data extraction: donepezil RCTs

Reference and design	Intervention	Participants	Outcome measures
Author: AD2000 Collaborative Group[43] Year: 2004 Country: UK Study design: Randomised, placebo-controlled, double-blind Number of centres: 22 hospitals Funding: Eisai and Pfizer sold the study drug and placebo to the researchers, but otherwise made no other contribution	Treatment arms: Patients took part in a run-in treatment period of 12 weeks, in which patients were randomly allocated either: (1) donepezil 5 mg/day or (2) placebo This was followed by a second randomisation to long-term donepezil: (1) donepezil 5 mg/day (n = 125) or 10 mg/day (n = 117) (2) placebo Other interventions used: Open-label aspirin treatment (enteric-coated 75 mg/day) or aspirin avoidance, continued uninterrupted unless a clear contraindication arose	Number of participants: 565 patients entered a 12-week run in period (1) donepezil (5 mg) (n = 282) (2) placebo (n = 283) 486 patients entered the second randomisation (1) donepezil (5 or 10 mg) (n = 242) (2) placebo (n = 244) Some patients who had previously received placebo were given donepezil. The numbers differ in the flow chart from in the text. The flow chart suggests that 9 patients who received placebo in the wash-in period received donepezil after the second randomisation. The text suggests that 125 patients who received placebo in the wash-in period received donepezil after the second randomisation, and that 116 patients who received donepezil in the wash-in period received placebo after the second randomisation Sample attrition/dropout: The number of dropouts from the donepezil and placebo groups was similar from week 13 onwards, with 152 of 183 (83%) allocated donepezil and 141 of 172 (82%) placebo who remained on study (i.e. not institutionalised, dead, or withdrawn), still taking AD2000 drugs at 60 weeks. This denominator excludes 59 from the donepezil group and 72 from the placebo group who were either institutionalised, dead or withdrawn. Proportions withdrawing including for these reasons, at 60 weeks will therefore be 152/242 (63%) for the donepezil group and 141/244 (58%) for the placebo group Sample crossovers: At the end of every 48 weeks phase of double-blind treatment period, a further 4-week treatment-free washout took place, whereupon patients could once again continue with another 48-week phase of double-blind treatment	Primary outcomes: Entry to institutional care (i.e. residential, nursing, or NHS continuing) and progression of disability, defined as loss of either two or four basic, or six of 11, instrumental activities on the Bristol activities of daily living score (BADLS) Secondary outcomes: Functional ability (BADLS: range 0–60); presence and severity of behavioural and psychological symptoms and signs of dementia, as measured by NPI (0–144); cognition measured with the MMSE (0–30); progress to severe cognitive disability (MMSE <10); psychological well-being of the principal caregiver, measured with the GHQ-30 (0–30); death from AD; safety; and compliance Also economic evaluation Methods of assessing outcomes: Clinical assessments were undertaken at baseline and 1–2 weeks before completion of each of the first five 12-week courses of donepezil or placebo. Assessments generally done by the local AD2000 nursing coordinator. Patients who completed the first 60 weeks of treatment were reassessed at the end of the 6-week treatment-free washout period. Assessments were repeated at week 114, and annually thereafter, for all patients, including those who had discontinued treatment. Patients opting to continue with AD2000 treatment had additional assessments after 12 weeks of every new phase, and after every 4-week washout period Clinicians completed a form describing any serious unexpected adverse events believed to be due to treatment

continued

Reference and design	Intervention	Participants	Outcome measures
		Inclusion/exclusion criteria for study entry: Patients were required to have a DSM IV diagnosis of dementia of Alzheimer's type, with or without a coexisting diagnosis of vascular dementia. Patients were required to have a regular carer, to be living in the community, and not to already be taking a cholinesterase inhibitor nor have a contraindication against donepezil. Doctor had to be substantially uncertain that the individual would obtain a worthwhile clinical benefit from donepezil, taking into account the available evidence and clinical circumstances Characteristics of participants: See below	Treatment compliance was validated through AD2000 pharmacy prescribing record cards and pill counts on returned treatment packs for patients who withdrew Mortality records were used to ensure long-term follow-up of survival and to obtain the certified cause of death Length of follow-up: originally intended to be 60 weeks but a protocol modification in 1999 gave option of indefinite extension. After a 6-week no-treatment washout, patients could continue with the same double-blind AD2000 treatment they had been receiving at 60 weeks for a further 48 weeks, if judged appropriate

Results

Patient characteristic at first randomisation	Donepezil (n = 283)	Placebo (n = 283)	p-Value
Dementia severity			
Mild (MMSE 19–26)	143 (51%)	148 (52%)	
Moderate (10–18)	149 (49%)	135 (48%)	
Men	118 (42%)	113 (40%)	
Age, years (median [range])	76 (54–93)	75 (46–90)	
Age group			
<60	8 (3%)	10 (4%)	
60–69	45 (16%)	49 (17%)	
70–79	163 (58%)	155 (55%)	
≥ 80	69 (24%)	69 (24%)	
Vascular dementia present	51 (18%)	42 (15%)	
Parkinsonism present	11 (4%)	11 (4%)	
Psychotic symptoms present	25 (9%)	29 (10%)	
Comorbidity present	149 (53%)	138 (49%)	
MMSE score (median [range])	19 (10–27)	19 (10–26)	
BADLS score (median [range])	13 (0–42)	15 (0–38)	
NPI score (median [range])	15 (0–84)	15 (0–74)	
GHQ-30 score median [range]	4 (0–27)	4.5 (0–29)	
Rates of institutionalisation (%)			
1 year	9	14	p = 0.15
3 years	42	44	p = 0.4

Comments: No significant difference was apparent between donepezil and placebo in rates of institutionalisation. The relative risk of entering institutional care, in the donepezil group compared with placebo was 0.97 (95% CI 0.72 to 1.30; p = 0.8). The number of patients institutionalised was slightly less with 10 mg than with 5 mg (37 vs 44, p = 0.7)

Time to loss of activities of daily living, institutional care, or both (%)			
1 year	13	19	p = 0.3
3 years	55	53	p = 0.9

Comments: Similar proportions of patients had progression of disability, and the relative risk of reaching endpoint, or entering institutional care, in the donepezil group compared with placebo was 0.96 (95% CI 0.74 to 1.24; p = 0.7). The relative risk of reaching disability endpoint alone was 1.02 (0.72–1.45; p = 0.9)

continued

	Donepezil	Placebo
Change from baseline in BADLS (number at risk) (*Estimated from graph*)		
Week 0	0 (n = 282)	0 (n = 283)
Week 12	–1 (n = 262)	–1 (n = 269)
Week 24	–2 (n = 220)	–3 (n = 230)
Week 36	–2.5 (n = 182)	–4 (n = 185)
Week 48	–3 (n = 162)	–5 (n = 162)
Week 60	–5 (n = 157)	–6.5 (n = 150)
Week 114?	–10.5 (n = 81)	–11.5 (n = 74)
Comments: No difference was apparent between donepezil and placebo on BADLS score at 12 weeks, but thereafter the donepezil group had better scores at all timepoints. The average difference was 1.0 BADLS points (0.5–1.6, p = 0.0004) better with donepezil than with placebo, with no significant rise or fall in efficacy over the first 2 years. BADLS scores were 1.0 points (95% CI –0.7 to 2.6, p = 0.24) better with 10 mg versus 5 mg of donepezil. Treatment effect 1.02 (SE 0.28) p < 0.0001. Patients who defaulted from treatment (retrieved dropouts) were 1.0 points worse, on average, than were those patients who remained on treatment (p = 0.009)		
Change from baseline in MMSE (number at risk) (*Estimated from graph*)		
Week 0	0 (n = 282)	0 (n = 283)
Week 12	1.0 (n = 245)	0 (n = 263)
Week 24	0.5 (n = 211)	0 (n = 229)
Week 36	0.4 (n = 185)	–1.0 (n = 192)
Week 48	0 (n = 165)	–1.5 (n = 168)
Week 60	–1.5 (n = 154)	–1.75 (n = 160)
Week 114	–5.0 (n = 94)	–5.0 (n = 87)
Comments: The donepezil group improved from baseline by an average of 0.9 MMSE points over the first 12 weeks, whereas no change was seen in the placebo group. Thereafter, both groups declined at similar rates. Over the 2-year study period, the donepezil group averaged MMSE scores of 0.8 points higher than the placebo group (95% CI 0.5 to 1.2, p < 0.0001) with no specific attrition of benefit. Cognition averaged 0.2 MMSE points (–0.8 to 1.2, p = 0.4) better with 10 mg than 5 mg of donepezil. No delay was found in reaching the severe cognitive disability milestone (MMSE <10), with a relative risk of 0.95 in the donepezil group compared with placebo (95% CI 0.64 to 1.41, p = 0.8). Treatment effect 0.83 (SE 0.18) p < 0.0001		
Change from baseline in NPI score (number at risk) (*Estimated from graph*)		
Week 0	0 (n = 282)	0 (n = 283)
Week 12	0 (n = 243)	2 (n = 260)
Week 24	1 (n = 209)	0 (n = 225)
Week 36	–1 (n = 180)	–2.5 (n = 186)
Week 48	0 (n = 160)	–3 (n = 162)
Week 60	–3 (n = 149)	–4 (n = 150)
Week 114	–4 (n = 81)	–7 (n = 71)
Comments: The difference between donepezil and placebo on behavioural and psychological symptoms was not significant at any timepoint, or overall, with the donepezil group averaging 0.3 points (95% CI –0.9 to 1.5, p = 0.6) better. These symptoms worsened by 1.7 NPI points less (95% CI –1.4 to 4.8, p = 0.3) with 10 mg compared with 5 mg. No selective benefit was noted among the 6% (41 of 565) of patients with NPI scores of 40 or more, indicating severe behavioural psychological symptoms (donepezil 1.7 worse than placebo compared with 0.4 better for those with NPI <40; p = 0.4). Treatment effect 0.31 (SE 0.59) p = 0.6		
Change from baseline in carer's GHQ score (number at risk) (*Estimated from graph*)		
Week 0	0 (n = 282)	0 (n = 283)
Week 12	1.0 (n = 246)	0.5 (n = 263)
Week 24	0.5 (n = 212)	0.5 (n = 229)
Week 36	0 (n = 182)	–0.2 (n = 192)
Week 48	–0.2 (n = 164)	–0.2 (n = 164)
Week 60	–0.5 (n = 151)	–0.5 (n = 153)
Week 114	–1.8 (n = 81)	–1.2 (n = 78)
Comments: Carers' psychological morbidity scores were 0.3 GHQ points (95% CI –0.3 to 0.9, p = 0.3) lower with donepezil compared with placebo. These scores were 0.8 points (–2.3 to 0.7, p = 0.3) worse with 10 mg donepezil than with 5 mg. Almost half of carers had scores of 5 or more at baseline, indicating probable psychological morbidity, and these proportions increased over time at about the same rate in both groups. Treatment effect 0.31 (SE 0.30) p = 0.3		

continued

	Donepezil	Placebo
Change from baseline for change in active care time (number at risk) (*Estimated from graph*)		
Week 0	0 (n = 282)	0 (n = 283)
Week 12	–0.1 (n = 243)	–0.1 (n = 260)
Week 24	–0.1 (n = 209)	–0.5 (n = 222)
Week 36	–0.3 (n = 179)	–0.7 (n = 184)
Week 48	–1.0 (n = 157)	–0.6 (n = 160)
Week 60	–1.0 (n = 147)	–1.7 (n = 140)
Week 114	–1.8 (n = 79)	–1.3 (n = 70)
Change from baseline for change in passive care time (*Estimated from graph*)		
Week 0	0 (n = 281)	0 (n = 282)
Week 12	0 (n = 242)	–2.0 (n = 260)
Week 24	–3.0 (n = 208)	–2.0 (n = 222)
Week 36	–4.0 (n = 179)	–4.0 (n = 184)
Week 48	–4.5 (n = 156)	–4.6 (n = 160)
Week 60	–6.0 (n = 148)	–6.0 (n = 139)
Week 114	–9.0 (n = 79)	–6.0 (n = 70)

Comments: Active caregiver daily input was 0.2 h less (95% CI –0.1 to 0.5, p = 0.2) and passive care time 0.4 h less (–0.5 to 1.2, p = 0.4) with donepezil. Change in active care time treatment effect: 0.19 (SE 0.14) p = 0.18. Change in passive care time: 0.37 (SE 0.41) p = 0.4

Adverse effects	Donepezil	Placebo	
No. of serious AEs[a]	29	23	p = 0.4
No. of deaths	63	50	p = 0.2
No. of patients withdrawn due to AE at week 12	36	20	p = 0.2

Comments: The number of serious adverse events and deaths were similar in those allocated donepezil or placebo. The underlying cause of 16 versus 11 (p = 0.7) of the deaths was certified as dementia

[a] Unclear over what time period this is calculated.

More patients allocated donepezil than placebo dropped out because of side effects and did not attend the 12-week assessment. No data presented for the 60-week or 114-week rates of dropout specifically due to side-effects.

Methodological comments

- Allocation to treatment groups: States telephone randomisation. No further information provided.
- Blinding: A numbered treatment box containing 12-week treatment packs of either donepezil or matching placebo was allocated and the first pack was obtained from the hospital pharmacy. Allocations (from available boxes) were minimised by age, severity of dementia, and presence or absence of vascular dementia, parkinsonism, and psychotic symptoms. States double-blind, no further details on blinding were provided.
- Comparability of treatment groups: Treatment groups for each comparison were closely balanced in terms of patient characteristics at entry.
- Method of data analysis: The 5 and 10 mg donepezil groups were combined unless otherwise stated. Standard logrank methods were used to compare rates of institutionalisation and progress of disability. Changes from baseline scores on the MMSE, BADLS, GHQ-30, NPI and caregiver time scales in the first 12 weeks and from 13 weeks onwards were analysed by the methods described by Fleiss (reference provided) and with multilevel models with repeated measures. Results were similar and only multilevel analyses are presented. The standard mixed model techniques included all randomised patients and allowed both for the crossover of treatment and for the fact that successive measurements at 12-week intervals from the same patient are correlated. By combining many different measurements into one estimate of overall treatment effect, the scale variables were analysed in the most efficient manner. Attempts were made to assess outcomes for all patients, including those who had discontinued. For the individual items with missing answers on questionnaires, the most recent previous score was used, if one existed. If not then the next subsequent valid score was substituted. Sensitivity analysis showed that results are not affected by this imputation versus insertion of a best or worst score, or by inflation of the total remaining score pro rata. The exception is the MMSE, for which the score for missing items was taken to be zero because the question was not successfully answered.

continued

- Sample size/power calculation: The target accrual was pragmatic, aiming to recruit, if possible, 3000 patients within 2–3 years, which would allow detection, or refutation, of any minimal improvements in the primary outcomes of institutionalisation and progress of disability, or the related cost-effectiveness measure. However, in response to slower than hoped for recruitment, the target was reduced, and the final sample size was 566, 482 of whom entered the long-term treatment. This number provides more than 90% power to detect, at $p < 0.05$, a 6-month delay in institutionalisation, which would avoid an average 2.6 weeks in care per-patient-year of treatment, given that 10% are institutionalised every year.
- Attrition/dropout: Dropouts reported at each phase of the study. Between 13 and 24 weeks, somewhat fewer individuals allocated donepezil than placebo were assessed (212 vs 227). This was a potential source of dropout bias because those who defaulted from treatment tended to have worse cognitive decline.

General comments
- Generalisability: Patients with mild to moderate Alzheimer's Disease with or without VAD – AD was the predominant disorder, (82% donepezil, 85% placebo) and therefore met our protocol.
- Outcome measures: Outcome measures were relevant to the study area and seem to have been measured appropriately.
- Inter-centre variability: none reported.
- Conflict of interests: Eisai and Pfizer sold the study drug and placebo but otherwise made no contribution to the study design or protocol, nor had any involvement in the management or reporting of the study.

Quality criteria for AD2000[43]

1. Was the assignment to the treatment groups really random?	Unknown
2. Was the treatment allocation concealed?	Unknown
3. Were the groups similar at baseline in terms of prognostic factors?	Reported
4. Were the eligibility criteria specified?	Adequate
5. Were outcome assessors blinded to the treatment allocation?	Inadequate
6. Was the care provider blinded?	Partial
7. Was the patient blinded?	Partial
8. Were the point estimates and measure of variability presented for the primary outcome measure?	Inadequate
9. Did the analyses include an intention-to-treat analysis?	Inadequate
10. Were withdrawals and dropouts completely described?	Adequate

Reference and design	Intervention	Participants	Outcome measures
Author: Burns *et al*.[50] Year: 1999 Country: International Study design: RCT double-blind, multicentre Number of centres: 82 Donepezil Study Group: Australia Belgium Canada France Germany Ireland New Zealand South Africa UK Funding: Eisai Inc., Teaneck, NJ, USA and Eisai Co. Ltd., Tokyo, Japan	Treatment arms: Patients were screened within 2 weeks of entry and randomly assigned to one of 3 study arms (1) 5 mg/day donepezil (2) 10 mg/day donepezil (3) Placebo Study medication was administered orally, once-daily, in the evening. For group 2 (10 mg per day donepezil) a blinded schedule was used where the patient initially received 5 mg/day for the first 7 days, then 10 mg/day for the remainder of the study A double-blind treatment phase of 24 weeks was followed by a 6 week single-blind placebo washout phase Other interventions used: None described	Number of participants: 818 randomised to treatment (1) 5 mg/day donepezil $n = 271$ (2) 10 mg/day donepezil $n = 273$ (3) Placebo $n = 274$ Sample attrition/dropout: A total of 187 patients discontinued (1) 5 mg/day donepezil $n = 60$ (22%) (2) 10 mg/day donepezil $n = 72$ (26%) (3) Placebo $n = 55$ (20%) Further details provided in results section Sample crossovers: No crossover Inclusion/exclusion criteria for study entry: Men and women 50 years or over with probable AD defined by DSM-III-R and NINCDS-ADRDA MMSE scores of 10–26 inclusive at screening and baseline Clinical Dementia Rating (CDR) scores of 1 (mild) or 2 (moderate) Computerised tomography or MRI scans within previous 6 months Women 2 years post-menopausal or surgically sterile Patients had to be generally healthy, with vision and hearing sufficient for compliance with testing procedures Written informed consent was obtained from both patients and caregivers Exclusion criteria: Patients with structural lesions or significant vascular changes Patients with other neurological or psychiatric disorders, asthma, significant uncontrolled gastrointestinal, renal, hepatic, endocrine or oncological disorders, or who were taking prohibited study medications, were all excluded Characteristics of participants: Age, years: mean ± SE (range) Gp1: 72 ± 0.5 (51–91); Gp2: 72 ± 0.5 (53–93); Gp3: 71 ± 0.5 (50–90) Gender, male/female (%male/%female): Gp1: 107/164 (39/61), Gp2: 118/155 (43/57), Gp3: 123/151 (45/55) Race, number (%): Caucasian: Gp1: 270 (100), Gp2: 271 (99), Gp3: 272 (99) Other: Gp1: 1 (<1), Gp2: 2 (1), Gp3: 2(1) Weight, kg: mean ± SE (range): Gp1: 65 ± 0.8 (38–108); Gp2: 66 ± 0.7 (38–99); Gp3: 66 ± 0.8 (37–107) Screening MMSE: mean ± SE (range): Gp1: 20 ± 0.3 (10–26); Gp2: 20 ± 0.2 (9–26); Gp3: 20 ± 0.3 (10–26)	Primary outcomes: ADAS-cog measure of cognition and CIBIC-plus measure of global functioning Secondary outcomes: Clinical Dementia Rating scale – sum of the boxes (CDR-SB) providing a consensus-based, global clinical measure from the 6 domains A modified Interview for Deterioration in Daily Living Activities in Dementia (IDDD), a measure of deterioration in activities of daily living in dementia Patient-rated QoL Adverse events Safety and physiological measures at each visit (not data extracted as per protocol) Primary outcomes: Efficacy and safety evaluations took place at baseline and at weeks 6, 12, 18, 24 and 30 Methods of assessing outcomes: ADAS-cog: no details (ref. given) CIBIC-plus: clinician-based interview with caregiver input (ref. given) CDR-SB: no details (ref. given) IDDD: initiation of tasks and their performance quantified by structured interview with the caregiver but no further details. To assess change in results of IDDD at subsequent visits the evaluator rated improvement, no change or deterioration in comparison to baseline performance but no details of who this evaluator was are given. QoL: patient rated but no other details (ref. given)

continued

Reference and design	Intervention	Participants	Outcome measures

Screening CDR: number (%) rating 0.5/1.0/.2.0: Gp1: 2(1)/222(82)/47(17); Gp2: 2(1)/236(86)/35(13); Gp3: 0/230(84)/44(16)

IDDD baseline severity scores – ITT population

	Donepezil		*Placebo*
	5 mg/day	*10 mg/day*	
IDDD-total score	67.78	69.85	69.84
Total baseline severity (possible range 33–231)	±1.61	±1.71	±1.68
Mean severity per item	2.04	2.09	2.10
Mean all patients	2.1		
IDDD-self care	23.14	23.72	24.31
Total baseline severity (possible range 16–112)	±0.61	±0.66	±0.72
Mean severity per item	1.45	1.48	1.52
Mean all patients	1.5		
IDDD-complex tasks	44.64 ±1.13	46.12 ±1.21	45.53 ±1.17
Total baseline severity (possible range 17–119)			
Mean severity per item	2.63	2.71	2.68
Mean all patients	2.7		

Means ± SE. The severity scale per item: 1–2 = no or slight impairment; 3–4 = mild impairment; 5–6 = moderate impairment; 7 = severe impairment giving a total score range of 33–231 points. The IDDD provides a self-care tasks rating (16 items) and a complex tasks rating (17 items)

ADAS-cog and CIBIC-plus baseline scores not given as presented as change from baseline. Assume similar from graph

Adverse events were monitored at each visit by questioning both the patient and the caregiver, as well as through direct observation. Analysis of adverse events was restricted to signs and symptoms that either began, or became more severe, after administration of the first dose of study medication. Events were coded using a modified COSTART dictionary and the assessment of relationship to treatment for all adverse events was conducted blind to treatment assignment. The incidence of adverse events and laboratory test abnormalities was compared between treatment groups using Fisher's exact test

All statistical analyses were undertaken by an independent clinical research organisation

Length of follow-up: 30 weeks: 24 weeks trial, 6 weeks single-blind placebo washout period

continued

Results

Outcomes		5 mg/day donepezil (n = 271)	10 mg/day donepezil (n = 273)	Placebo (n = 274)	Other
ADAS-cog LS mean change from baseline ± SE Values in italics estimated from Fig. 1	Week 6	*–1.8 ± 0.3* $p = 0.0005$*	*–2.1 ± 0.3* $p < 0.0001$*	*–0.3 ± 0.3*	* Assume from graph that p-values relate to change from placebo, not change from baseline; however, not made clear in text except for general statement that there was significant improvement compared to placebo
	Week 12	*–1.6 ± 0.3* $p < 0.0001$*	*–1.9 ± 0.3* $p < 0.0001$*	*0.4 ± 0.3*	
	Week18	*–0.6 ± 0.4* $p = 0.0002$*	*–1.7 ± 0.4* $p < 0.0001$*	*1.3 ± 0.4*	
	Week 24	*0.4 ± 0.4* $p = 0.0315$*	*–1.4 ± 0.4* $p < 0.0001$*	*1.5 ± 0.4*	
	Endpoint	*0.2 ± 0.3* $p = 0.0021$*	*–1.2 ± 0.3* $p < 0.0001$*	*1.7 ± 0.3*	Donepezil vs placebo differences of 1.5 and 2.9 points for 5 and 10 mg/day donepezil groups respectively
	Week 30	*3.4 ± 0.4*	*2.5 ± 0.5*	*2.9 ± 0.4*	

Comments: A negative mean change in ADAS-cog indicates a clinical improvement. Results at endpoint are ITT analysis using LOCF technique at 24 weeks. As this differs from the 24 week endpoint the other results may not be ITT. The paper does not make this clear. Values estimated from graphs for the endpoint: these values give donepezil vs placebo differences that are in agreement with values in the text. There was a statistically significant improvement in ADAS-cog scores for the two donepezil-treated groups compared with placebo. This was observed at Week 6 and maintained throughout the active treatment phase

Outcomes		5 mg/day donepezil (n = 271)	10 mg/day donepezil (n = 273)	Placebo (n = 274)	Other
CIBIC plus scores Mean ± SE p-values calculated by Cochran–Mantel–Haenszel test Values in italics estimated from Fig. 2	Week 6	*3.84 ± 0.05* $p = 0.0029$	*3.87 ± 0.06* $p = 0.0023$	*4.05 ± 0.05*	All p-values: assume from graph that p-values relate to change from placebo, not change from baseline; however, not made clear in text except for general statement that there was significant improvement compared to placebo
	Week 12	*4.03 ± 0.06* $p < 0.0545$	*3.90 ± 0.05* $p = 0.0001$	*4.23 ± 0.06*	
	Week18	*4.08 ± 0.06* $p = 0.0010$	*4.00 ± 0.06* $p < 0.0001$	*4.45 ± 0.06*	
	Week 24	*4.23 ± 0.8* $p = 0.0326$	*4.08 ± 0.08* $p = 0.0009$	*4.49 ± 0.06*	
	Endpoint	*4.23 ± 0.06* $p = 0.0072$	*4.13 ± 0.06* $p = 0.0002$	*4.52 ± 0.06*	
	Week 30	*4.57 ± 0.9*	*4.57 ± 0.09*	*4.78±0.08*	

Comments: A value of 4 was the boundary between clinical improvement or decline, a value less than 4 indicates clinical improvement, a value greater than 4 clinical decline. Results at endpoint are ITT analysis using LOCF technique at 24 weeks. As this differs from the 24-week endpoint the other results may not be ITT. The paper does not make this clear. Statistically significantly greater numbers of donepezil-treated patients were judged clinically improved, by comparison with placebo. This beneficial drug treatment effect was observed from Week 6 and was maintained at all subsequent visits and at endpoint

continued

		5 mg/day donepezil (*n* = 271)	10 mg/day donepezil (*n* = 273)	Placebo *n* = 274	
% Patients rated as improved (CIBIC-plus scores ≤ 3 at endpoint)		21	25	14	
% Treatment failures (CIBIC-plus scores ≥ 5 at endpoint)		43	37	51	
CDR-SB scores LS mean change from baseline ± SE Values in italics estimated from Fig. 3	Week 6	*–0.11 ± 0.05* no *p*-value given	*–0.11 ± 0.06* no *p*-value given	*–0.02 ± 0.06*	
	Week 12	*–0.18 ± 0.08* $p = 0.0021$	*–0.18 ± 0.08* $p = 0.0014$	*0.15 ± 0.08*	
	Week 18	*–0.02 ± 0.09* $p = 0.0154$	*–0.17 ± 0.11* $p = 0.0006$	*0.33 ± 0.11*	
	Week 24	*0.06 ± 0.12* $p = 0.0387$	*–0.11 ± 0.12* $p = 0.0020$	*0.39 ± 0.12*	
	Endpoint	*0.06 ± 0.11* $p = 0.0344$	*–0.06 ± 0.11* $p = 0.0033$	*0.37 ± 0.06*	Donepezil vs placebo differences of 0.3 and 0.4 points for 5 and 10 mg/day donepezil groups respectively
	Week 30	*0.55 ± 0.14*	*0.49 ± 0.15*	*0.68 ± 0.14*	

Comments: A positive score indicates clinical decline, a negative score indicates clinical improvement. Results at endpoint are ITT analysis using LOCF technique at 24 weeks. As this differs from the 24 week endpoint the other results may not be ITT. The paper does not make this clear. Statistically significant improvements in LS mean change CDR-SB scores were observed for both donepezil-treated groups, versus placebo, at weeks 12, 18, 24 and endpoint ($p < 0.05$)

Outcomes		5 mg/day donepezil (*n* = 271)	10 mg/day donepezil (*n* = 273)	Placebo (*n* = 274)	Other
IDDD-Complex Task scores LS mean change from baseline (±SE) Values in italics estimated from Fig. 4	Week 6	*67.7*** $p = 0.0155$	*67.6*** $p = 0.0065$	*68.8 ± 0.4*	**SE not estimable from graph
	Week 12	*69.0 ± 0.4* *p* not given, nsd	*68.0 ± 0.4* $p = 0.0085$	*69.4 ± 0.4*	
	Week 18	*69.6 ± 0.4* *p* not given, nsd	*68.7 ± 0.4* $p = 0.0033$	*70.5 ± 0.4*	
	Week 24	*70.8 ± 0.6* *p* not given, nsd	*69.2 ± 0.6* $p = 0.0163$	*71.0 ± 0.5*	
	Endpoint	70.4 ± 0.4 *p* not given, nsd	*69.4 ± 0.4* $p = 0.0072$	*71.1 ± 0.4*	
	Week 30	*72.5***	*73.1 ± 0.6*	*71.8 ± 0.6*	**SE not estimable from graph

Comments: A value of 68 was the boundary between clinical improvement or decline, a value less than 68 indicates clinical improvement, a value greater than 68 clinical decline. Results at endpoint are ITT analysis using LOCF technique at 24 weeks. As this differs from the 24 week endpoint the other results may not be ITT. The paper does not make this clear. No improvements in IDDD-self care could be measured in the study because this patient population was not impaired at baseline. See above for descriptions of severity. From week 6, through the active treatment phase, IDDD-complex task scores for both the 5 and 10 mg/day donepezil groups were improved when compared with placebo, with statistical significance for the 10 mg/day donepezil dose at all assessments. In assessment of change from baseline, a 7-point Likert-type scale was used, where 1 = marked improvement from baseline, 4 = no change from baseline and 7 = marked deterioration from baseline

continued

QoL Not presented in detail in the paper because the mean change from baseline at each evaluation of this patient rated measure was associated with a large standard error indicating the high variability of responses from patients. No clear trends among the treatment groups were evident

General comments: Authors highlight that following the 6-week, single-blind, placebo washout phase patient scores for efficacy measures reverted to levels similar to placebo, indicating that the beneficial effects of donepezil were lost when treatment was discontinued

Completion rate (%)	**78**	**74**	**80**	**76**
	5 mg/day donepezil (*n* = 271)	**10 mg/day donepezil (*n* = 273)**	**Placebo (*n* = 274)**	**All donepezil treatments *n* = 544**
Patients discontinued (%)	60 (22)	72 (26)	55 (20)	132 (24)
Withdrawn due to:				
Adverse Events**	24 (9)	50 (18)	27 (10)	74 (14)
Body as a whole	4 (1)	12 (4)	6 (2)	16 (3)
Cardiovascular	1 (<1)	5 (2)	3 (1)	6 (1)
Digestive	4 (1)	27 (10)	2 (<1)	31 (6)
Nervous	13 (5)	21 (10)	14 (5)	34 (6)
Intercurrent Illness	0	0	3 (1)	0
Request of patient or investigator	12 (4)	6 (2)	6 (2)	18 (3)
Non-compliance	3 (1)	2 (1)	2 (1)	5 (1)
Protocol violation	13 (5)	8 (3)	13 (5)	21 (4)
Other	8 (3)	6 (2)	4 (1)	14 (3)

Comments: ** **There may be more than one adverse event that led to withdrawal.** Adverse events were not necessarily treatment related or treatment emergent

Adverse events experienced by at least 5% of all donepezil patients	**5 mg/day donepezil (*n* = 271)**	**10 mg/day donepezil (*n* = 273)**	**Placebo (*n* = 274)**	**All donepezil treatments *n* = 544**
Total patients with any adverse event (%)	213 (79)	234 (86)	207 (76)	447 (82)
Digestive system (%)**	70 (26)	127 (47)	65 (24)	197 (36)
Nausea**	7%	24%	7%	16%
Diarrhoea**	10%	16%	4%	13%
Vomiting**	4%	16%	4%	10%
Anorexia	4%	8%	1%	6%
Nervous system (%)**	98 (36)	109 (40)	80 (29)	207 (38)
Dizziness	5%	9%	5%	7%
Confusion	7%	6%	6%	7%
Insomnia	7%	8%	4%	8%
Total patients with serious adverse events (SAE) (%)	19 (7)	29 (11)	25 (9)	73 (9)

Comments: ** Donepezil groups significantly differed from placebo, $p \leq 0.05$ employing Fisher's exact test

All adverse events, whether reported or observed, were recorded together with the time and date of onset and cessation, severity of condition and whether, in the opinion of the investigator, the event was related to donepezil treatment. Serious adverse events (SAE) included fatal or life-threatening situations, permanently disabling conditions or incidents that required or prolonged hospitalisation

The most frequently experienced adverse events were digestive system related which are predictable effects of cholinergic drugs. The majority of events were mild and transient, typically lasting 1–2 days and resolving during continued donepezil use, without dosage adjustment. Most adverse events, other than those clearly cholinergic in nature, were judged by the investigators not to be related to donepezil treatment

continued

Deaths during the study or within one month of stopping medication	5 mg/day donepezil (*n* = 271)	10 mg/day donepezil (*n* = 273)	Placebo (*n* = 274)
	1	2	2

Comments: All five deaths were determined to be unrelated to donepezil treatment.

Methodological comments

- Allocation to treatment groups: Paper states patients were randomised to treatment but no details of how this was carried out are given. No details of allocation concealment.
- Blinding: Paper states study is double-blinded with a single-blind placebo washout phase following treatment. No details regarding the methods used to achieve blinding to treatment are given. Some outcome measures were clinician assessed and some were patient or caregiver assessed. It is unclear whether assessors were unaware of the treatment allocation for all assessments. It is unclear whether the clinical assessor was the same person as the treating physician. An independent clinical research organisation undertook the statistical analyses but in general it is unclear whether these analysts were blind to the treatment groups. Exceptions are the initial examination of the ADAS-cog data which was blinded and the assessment of relationship to treatment for all adverse events.
- Comparability of treatment groups: Treatment groups appear to be comparable on baseline psychological measures. Reports that all groups comparable with respect to demographic variables examined.
- Method of data analysis: Three populations were used in the analyses of efficacy: fully evaluable, retrieved dropout and intention-to-treat (ITT). The ITT population was analysed on both observed cases and traditional last observation carried forward (LOCF). As specified *a priori*, the primary population was the ITT and the primary endpoint was the week 24 LOCF. Since the results of all analyses were similar, only the primary analysis is presented in this report. Graphs give week 6, week 12, week 18, week 24 and week 30 data in addition to an outcome labelled endpoint. The endpoint is the ITT (LOCF) week 24 result, but week 6, week 12, week 18, week 24 and week 30 are not ITT, although this is not made clear in the paper. For continuous variables (ADAS-cog, modified IDDD, CDR-SB and QoL) an analysis of covariance model was used to compare treatment groups. For CIBIC-plus the Cochran–Mantel–Haenszel test was employed, with RIDITS as the score option (the meaning of this is not explained in the text) and stratified for centre. Demographic variables (age, weight, height) were investigated with ANOVA models with factors for treatment and centre. Sex was assessed by the Cochran–Mantel–Haenszel test with centres as strata. Between-group differences were investigated by ANCOVA models. Fisher's least significant difference procedure was used to control for multiple comparisons to placebo. Incidence of adverse events was compared between treatment groups using Fisher's exact test. All hypothesis tests were two-sided and statistical significance was achieved if $p \leq 0.05$. Most data was reported as mean $\pm$SE, *p*-values were given.
- Sample size/power calculation: An original sample size of 150 patients/treatment group was estimated based on the results obtained from an earlier Phase II study of donepezil and from published results from US tacrine trials. The sample size had 80% power to detect a difference of 0.27 points in mean CIBIC-plus scores for each donepezil treatment group when compared to placebo, at a 0.05 significance level. The patient completion rate was estimated to be 80%. During the study, a blinded examination of the ADAS-cog test data indicated a larger variance in the multinational data than projected from the US data. This was not unexpected given the multinational nature of the patient cohort. However, to ensure a valid representation of the patient cohort it was thus necessary that the sample size be increased to 250 patients per treatment group. The final sample size of 818 was a result of additional patients already in screening at the time of termination of recruitment.
- Attrition/dropout: Dropouts were recorded and the reasons for these were given. This information was given for the study as a whole so there is no way of telling when the dropouts occurred and thus what the loss was at each time point where outcome measures were assessed. As the data presented for the time points may not have been calculated on an ITT basis it would have been helpful to know how many patients contributed to those results. Completion rates were: Gp1: 78% Gp2: 74% Gp3: 80% All donepezil treatments: 76%. Dropout rates were: Gp1: 60/271 (22%) Gp2: 72/273 (26%) Gp3: 55/274 (20%) All donepezil treatments: 132/544 (24%).

General comments

- Generalisability: The patients included in this study did not have concurrent diseases of the following types: neurological or psychiatric disorders, asthma, significant gastrointestinal, renal, hepatic, endocrine or oncological disorders. Therefore the results may not be applicable to patients who do have such concurrent diseases. The taking of 'prohibited study medications' excluded patients from participating but these medications are not further described. Applicable to patients with mild to moderate AD (MMSE score 10–26 inclusive) and with CDR scores of 1 (mild) or 2 (moderate).
- Outcome measures: Appear to be relevant to study area. One, the IDDD was modified to assess change from baseline, this trial represents the first use of this modified scale. There is no mention of whether compliance was assessed in anyway.
- Inter-centre variability: Not stated in the paper.
- Conflict of interests: The research was funded by Eisai Inc (Teaneck, NJ, USA) and Eisai Co. Ltd. (Tokyo, Japan); one author works for the sponsor.

Quality criteria for Burns *et al.*[50]

1. Was the assignment to the treatment groups really random?	Unknown
2. Was the treatment allocation concealed?	Unknown
3. Were the groups similar at baseline in terms of prognostic factors?	Reported
4. Were the eligibility criteria specified?	Adequate
5. Were outcome assessors blinded to the treatment allocation?	Partial
6. Was the care provider blinded?	Partial
7. Was the patient blinded?	Partial
8. Were the point estimates and measure of variability presented for the primary outcome measure?	Adequate
9. Did the analyses include an intention-to-treat analysis?	Adequate
10. Were withdrawals and dropouts completely described?	Adequate

Reference and design	Intervention	Participants	Outcome measures
Author: Gauthier *et al.*[41] Year: 2002 Country: multinational Study design: RCT, double-blind multicentre (Canada, Australia, France) **NB substudy of moderate patients (MMSE 10–17) from Feldman[42]** Number of centres: 3 Funding: Pfizer Inc and Eisai Inc	Treatment arms: (1) Donepezil 5 mg/day for 28 days (single dosage) followed by an increase to 10 mg/day. Study medication could be reduced to 5 mg/day at any time to improve tolerability (2) Placebo 79% of donepezil patients and 90% placebo patients attained maximum daily dosage of 10 mg of donepezil/placebo equivalent. Of these, 8 donepezil and 3 placebo patients subsequently had their dosage lowered to 5 mg/day	Number of participants: n = 207 (102 donepezil patients and 105 placebo patients) Sample attrition/dropout: Dropout rates were 19% in the donepezil group and 11% in the placebo group. The difference was not significant. The most common reasons for discontinuation were adverse events (9% donepezil, 5% placebo) and withdrawal of consent (3% donepezil, 5% placebo) Sample crossovers: None Inclusion/exclusion criteria for study entry: Subanalysis patients met criteria for AD and had MMSE scores of 10–17. Patients in the community or in assisted living facilities were included so long as they didn't require total nursing care. Patients confined to wheelchairs were excluded Characteristics of participants: Donepezil (n = 102)/placebo (n = 105) Mean age (range) 74.3 (52–92)/74.3 (48–90) No. (%) women 70 (68.6)/60 (57.1) LS mean baseline scores ± SE (range) CIBIS 4.12 ± 0.07 (3.0–6.0)/4.16 ± 0.06 (3.0–6.0) sMMSE 13.57 ± 0.29 (10.0–17.0)/13.86 ± 0.26 (10.0–18.0*) SIB 85.65 ± 1.12 (38.0–99.0)/85.27 ± 1.07 (50.0–99.0) DAD 58.00 ± 2.33 (5.0–94.4)/58.31 ± 2.16 (2.7–100.0)	Primary outcomes: CIBIC-plus Secondary outcomes: Two cognitive scales – sMMSE; Severe Impairment Battery (SIB) and 3 functional assessments: Disability Assessment for Dementia (DAD); modified Instrumental Activities of Daily Living (IADL+) and Physical Self-Maintenance Scale (PSMS+). Behavioural and neuropsychiatric symptoms were measured using the 12-item Neuropsychiatric Inventory (NPI). Patient assessments were carried out at weeks 4, 8, 12, 18 and 24, except for the sMMSE and DAD (weeks 12 and 24 only) and the IADL+ and PSMS+ (weeks 4, 12 and 24). Safety was evaluated by means of medical history, physical examinations, vital signs, clinical laboratory tests and ECGs, plus monitoring of adverse events. Only adverse events were data extracted. Methods of assessing outcomes: not stated Length of follow-up: 24 weeks

continued

Reference and design	Intervention	Participants	Outcome measures
	Other interventions used: Most concomitant medicines were allowed, except those with notable cholinomimetic or anticholinergic effects, and investigational drugs	IADL+ 59.71 ± 2.50 (12.5–97.9)/59.32 ± 2.30 (5.6–100.0) PSMS+ 6.68 ± 0.41 (3.0–20.0)/6.73 ± 0.38 (3.0–17.0) NPI 12-item total 17.87 ± 1.69 (0.0–76.0)/ 16.74 ± 1.60 (0.0–86.0) * One patient with a baseline sMMSE score of 18 was randomised into the trial The higher proportion of women in the donepezil group was not significant	

Results

CIBIC-plus least-squares mean scores	Donepezil	Placebo	*p*-Value
Week 4 (donepezil = 94, placebo = 97)	3.8	4.0	
Week 8 (donepezil = 79, placebo = 85)	3.6	3.95	0.0068
Week 12 (donepezil = 86, placebo = 96)	3.55	4.05	0.0002
Week 18 (donepezil = 83, placebo = 93)	3.75	4.2	0.0032
Week 24 (donepezil = 83, placebo = 93)	3.95	4.4	0.0044
Week 24 LOCF (donepezil = 98, placebo = 105)	4.0	4.5	0.0003

Comments: Numbers were all estimated from figure. A score of 4 indicates no change, <4 indicates clinical improvement, >4 clinical decline. At week 24 LOCF, 70% of donepezil patients and 47% of placebo patients were rated as improved or no change ($p = 0.0007$)

sMMSE least-squares mean change from baseline scores	Donepezil	Placebo	*p*-Value
Week 12 (donepezil = 84, placebo = 96)	2	0	0.0004
Week 24 (donepezil = 83, placebo = 91	1.6	–0.4	0.0009
Week 24 LOCF (donepezil = 91, placebo = 100)	1.5	–0.56	0.0002

Comments: Numbers were all estimated from figure. Positive scores indicate clinical improvement. Week 24 LOCF mean treatment difference = 2.06

SIB least-squares mean change from baseline scores	Donepezil	Placebo	*p*-Value
Week 4 (donepezil = 93, placebo = 99)	1.7	0.5	
Week 8 (donepezil = 79, placebo = 85)	3	0.4	0.0066
Week 12 (donepezil = 85, placebo = 95)	3.5	–0.4	0.0004
Week 18 (donepezil = 83, placebo = 93)	3.9	–1.0	0.0002
Week 24 (donepezil = 83, placebo = 93)	2.5	–3.0	0.0012
Week 24 LOCF (donepezil = 98, placebo = 104)	1.4	–3.0	0.0026

Comments: Numbers were all estimated from figure. Positive scores indicate clinical improvement. Week 24 LOCF mean treatment difference = –4.44

DAD least-squares mean change from baseline scores.	Donepezil	Placebo	*p*-Value
Week 12 (donepezil = 86, placebo = 96)	2.5	–4	0.0037
Week 24 (donepezil = 83, placebo = 93)	0.5	–9.0	<0.0001
Week 24 LOCF (donepezil = 92, placebo = 101)	0.0	–9.25	<0.0001

Comments: Numbers were all estimated from figure. Positive scores indicate clinical improvement. Week 24 LOCF mean treatment difference = –9.25

continued

NPI least-squares mean change from baseline scores	Donepezil	Placebo	*p*-Value
Week 4 (donepezil = 92, placebo = 99)	–4	–0.7	0.0387
Week 8 (donepezil = 78, placebo = 85)	–3.5	–1.0	–
Week 12 (donepezil = 78, placebo = 85)	–3.6	–1.0	–
Week 18 (donepezil = 85, placebo = 95)	–3.4	–0.1	–
Week 24 (donepezil = 81, placebo = 93)	–5.0	–0.8	0.021
Week 24 LOCF (donepezil = 97, placebo = 107)	–5.0	0.92	0.0022

Comments: Numbers were all estimated from figure. Negative scores indicate clinical improvement. NPI at week 24 LOCF mean treatment difference = 5.92. On the IADL+ and PSMS+ there were significant differences in favour of donepezil treatment compared with placebo at week 24 and week 24 LOCF (mean treatment difference =7.81, $p = 0.0002$ on the IADL+ and 1.31, $p = 0.001$ on the PSMS+). Individual NPI item analysis at week 24 LOCF showed benefit with donepezil compared with placebo on all 12 items of the NPI, with significant differences for delusions ($p = 0.0073$), apathy ($p = 0.0131$) and aberrant motor behaviour ($p = 0.0232$)

Adverse events occurring in ≥5% of patients receiving donepezil *n* (%)	Donepezil (*n* = 102)	Placebo (*n* = 107)
Any adverse event	84 (82.4)	84 (80.0)
Diarrhoea	13 (12.7)	6 (5.7)
Headache	11 (10.8)	4 (3.8)
Respiratory tract infection	11 (10.8)	11 (10.5)
Asthenia	10 (9.8)	5 (4.8)
Arthralgia	9 (8.8)	2 (1.9)
Nausea	8 (7.8)	4 (3.8)
Back pain	8 (7.8)	6 (5.7)
Dizziness	8 (7.8)	4 (3.8)
Weight loss	8 (7.8)	4 (3.8)
Vomiting	7 (6.9)	3 (2.9)
Accidental injury	7 (6.9)	10 (9.5)
Abdominal pain	7 (6.9)	8 (7.6)
Hostility	6 (5.9)	7 (6.7)
Dyspepsia	6 (5.9)	2 (1.0)
Urinary tract infection	6 (5.9)	4 (3.8)

Comments: The majority or AEs (66%) were rated as mild in severity and, in general, were similar between the two groups. Moderate (28%) or severe (6%) AEs were also similarly distributed between the two groups. AEs that are predominantly cholinergic in nature (i.e. nausea, vomiting and diarrhoea) from the current study ($n = 207$, dose increase after 4 weeks at 5 mg/day) show a lower incidence in comparison with placebo than earlier studies where the dose was increased after only one week. A total of 27 patients (14% donepezil, 12% placebo) experienced serious adverse events. In the donepezil patients, all of the serious AEs were considered unrelated to donepezil by the investigator. 10% of donepezil patients and 5% of placebo patients withdrew due to AEs.

Methodological comments

- Allocation to treatment groups: Patients were part of a larger study, therefore randomisation was based on a larger sample. In the previous study it was reported that randomisation was by computer schedule.
- Blinding: Main paper states that identical tablets were used.
- Comparability of treatment groups: There was no imbalance in demographic characteristics including age, gender or race. There were no differences in baseline outcome measures. The difference in % of female patients was not statistically significant [donepezil: 70 (68.6%), placebo: 60 (57.1%)].
- Method of data analysis: Primary analysis of efficacy was based on the change from baseline scores on the CIBIC-plus at week 24 in the ITT population using LOCF where there were missing values. ITT population consisted of all randomised patients who took at least one dose of study medication and provided a baseline assessment and at least one post-baseline efficacy assessment. Secondary analyses were carried out using observed case analysis at each visit. All statistical tests were two sided, and *p*-values of ≤0.05 were considered statistically significant. No adjustments were carried out for multiple endpoint comparisons. Graphs showed means scores ±SE. Some patients were missing from the analysis, therefore ITT not to the true definition.
- Sample size/power calculation: This is a substudy of another trial. The main trial was powered for 96 patients in each group, and this substudy has >100 patients in each group, so it is assumed that the present study is adequately powered.
- Attrition/dropout: Dropout rates stated and most common two reasons given. Other reasons not stated.

continued

General comments
- Generalisability: The study was a subgroup of moderate-severe patients from a previous study. This previous study was excluded on the bases of the participant group.
- Outcome measures: Appropriate measures were used.
- Inter-centre variability: Not discussed.
- Conflict of interests: Two of the authors work for Pfizer Pharmaceuticals Inc, and the study was supported by Pfizer Inc and Eisai Inc.

Quality criteria for Gauthier *et al.*[41]

1. Was the assignment to the treatment groups really random?	Unknown
2. Was the treatment allocation concealed?	Unknown
3. Were the groups similar at baseline in terms of prognostic factors?	Reported
4. Were the eligibility criteria specified?	Partial
5. Were outcome assessors blinded to the treatment allocation?	Unknown
6. Was the care provider blinded?	Unknown
7. Was the patient blinded?	Unknown
8. Were the point estimates and measure of variability presented for the primary outcome measure?	Inadequate
9. Did the analyses include an intention-to-treat analysis?	Inadequate
10. Were withdrawals and dropouts completely described?	Partial

Gauthier *et al.*[41] is a substudy investigating the use of donepezil for patients with moderate AD (MMSE score 10–17). The original study by Feldman *et al.* (2001) investigated the use of donepezil in patients with moderate to severe AD and was therefore excluded from this review. However, results from the Feldman *et al.* study showed that patients receiving donepezil showed benefits on the CIBIC+, compared with placebo, at all visits up to week 24 ($p < 0.001$) and at week 24 LOCF ($p < 0.0001$). All secondary measures (including sMMSE, SIB, DAD, FRS and NPI) showed significant differences between the groups in favour of donepezil at week 24 LOCF. These data suggest that donepezil's benefits extend into more advanced stages of AD than those previously investigated, with very good tolerability.

Reference and design	Intervention	Participants	Outcome measures
Author: Greenberg *et al*.[54] Year: 2000 Country: USA Study design: Two-centre, randomised, placebo-controlled, double-blind, crossover study Number of centres: Two Funding: National Institute of Aging (Bethesda, MD) through the Massachusetts Alzheimer's Disease Research Center and Massachusetts General Hospital, Mallinckrodt General Clinical Research Center	Treatment arms: This was a 24-week study protocol. All individuals began with a six-week single-blind placebo wash in period. Patients were then randomised with equal allocation to one of two treatment schedules A or B (double masked). Group A: received six weeks of placebo 'treatment' and then crossed over to donepezil 5 mg/day for the next six weeks. A six-week placebo washout made up the final six weeks of the trial period. Group B: received six weeks of donepezil 5 mg/day and then crossed over to a six-week washout with placebo. A six-week placebo 'treatment' made up the final six weeks of the trial period Study medication was taken orally, once daily in the evening for six weeks Other interventions used: 12 participants were taking concomitant psychoactive medication: 8 taking anti depressants, 2 taking sedatives and 2 taking both antidepressants and sedatives	Number of participants: A total of 64 patients entered the wash-in phase of the study, 60 of whom were randomised to a crossover sequence, 30 in group A and 30 in group B Sample attrition/dropout: Four patients withdrew before randomisation. Nine patients withdrew after randomisation, 5 in group A (2 during placebo therapy, 3 during donepezil therapy) and 4 in group B (3 during donepezil therapy and 1 during placebo therapy). See results for reasons Sample crossovers: There were 30 participants group A and group B. By the crossover point of the study 28 participants remained in group A and 27 participants in group B (calculated by reviewer) Inclusion/exclusion criteria for study entry: Men and women with a diagnosis of probable AD. Ability to undergo cognitive testing (defined as an information-memory-concentration subscale score of 20 or less) – this requirement restricted the study population largely to those with mild to moderate dementia. Six years or more of education, fluency in speaking English, stable doses of any concomitant medication for 4 weeks before enrolment and the presence of an appropriate caregiver to monitor medication use and attend all follow-up assessments. Exclusion criteria: Specific contraindications to cholinesterase inhibitor use such as a history of sick sinus syndrome or other supraventricular conduction defect. Active gastrointestinal tract bleeding. Bladder obstruction. Asthma or severe obstructive pulmonary disease. Hypersensitivity to cholinesterase inhibitor use. Having taken cholinesterase inhibitors within the previous 3 months Written informed consent was obtained from patients and family caregivers Characteristics of participants: Group A = placebo/donepezil $n = 30$ Group B = donepezil/placebo $n = 30$ Sex M/F GpA 18/12; GpB 18/12; Total 30/30 Note this doesn't add up: Either the total is M/F 36/24 or one of the other groups is M/F 12/18. The following are Mean ± SD Age in years GpA 74.9 ± 10.1; GpB 75.1 ± 9.0; Total 75.0 ± 9.5	Primary outcomes: ADAS-cog Secondary outcomes: Explicit verbal recall (assessed by NYU Stories Test, delayed recognition subscale) Verbal fluency Caregiver-rated global impression of change: These interviews took place at weeks 6, 12, 18 and 24. Caregivers were asked to rate patient function in the previous six weeks as mildly to markedly worsened, unchanged, or mildly to markedly improved. Interview with caregiver to verify concurrent medication use and adverse events Methods of assessing outcomes: Cognitive testing was performed by a psychometrician. A separate set of personnel scored the cognitive tests and entered data into a database. Who interviewed the caregivers is not stated Evaluations were performed at 6, 12, 15, 18, 21 and 24 weeks. This schedule ensured testing took place at the beginning and end of their donepezil and placebo treatments and after 3 weeks of drug washout Length of follow-up: 24 weeks, 6 weeks of run-in followed by 18 weeks of treatment, washout and a second treatment period

continued

Reference and design	Intervention	Participants	Outcome measures
	Whether any other participants were taking other medications is not described	Disease duration in years GpA 4.1 ± 2.8; GpB 3.5 ± 2.1; Total 3.8 ± 2.5 Education in years GpA 15.2 ± 3.5; GpB 14.2 ± 3.5; Total 14.7 ± 3.5 MMSE Score GpA 21.6 ± 3.5; GpB 21.9 ± 4.0; Total 21.8 ± 3.7 BDS Score (Blessed Dementia Scale information-memory concentration subscale) GpA 11.1 ± 4.1; GpB 10.0 ± 4.4; Total 10.5 ± 4.2 ADAS-cog Score GpA 18.7 ± 7.4; GpB 18.3 ± 8.2; Total 18.5 ± 7.7	

Results: Results in this paper were presented in several different ways, for Groups A and B, for all who received placebo (whether in A or B) and similarly for all who received donepezil, for those who completed both parts of the crossover and for whom no data points were missing ($n = 48$ of a possible 60). Caution required

Outcomes	**Group A: placebo then donepezil ($n = 30$ at start)**	**Group B: donepezil then placebo ($n = 30$ at start)**	***p*-Value**
Completed treatment	25	26	
	Improved during donepezil but not placebo treatment	**Improved during placebo but not donepezil treatment**	
Patients improved (ADAS-cog scores decreased)	21 (44%) of 48 patients	9 (19%) of 48 patients	$p = 0.03$

Comments: Combining within-individual changes during drug and placebo use, ADAS-cog scores showed a 2.17-point (95% CI 0.20–4.10 points) net improvement in response to donepezil administration

	Placebo treatment $n = 52$	**Donepezil treatment $n = 51$**	**Donepezil–placebo both treatments Difference between changes in ADAS-cog scores during donepezil and placebo treatment. (Negative score indicates improvement) $n = 48$**
Change in ADAS-cog Mean±SEM	+0.62 ± 0.61	–1.50 ± 0.58 $p < 0.05$	–2.17 ± 0.98 $p < 0.05$

Comments: ADAS-cog: maximal impairment is a score of 70, lower scores indicate less severity. Scores on the ADAS-cog could not be determined for every randomised patient because of study dropout ($n = 9$) or inability to complete ADAS-cog testing at a particular visit. The number of patients able to complete all testing with donepezil and placebo was 48. There was no effect on response to donepezil therapy associated with patient age, sex, level of education, disease duration, centre of enrolment or severity of dementia at baseline

continued

Washout of drug effect After donepezil treatment the mean change in ADAS-cog score as given above was –1.50 ± 0.58 ($n = 51$) $p < 0.05$. After 3 weeks of placebo washout this change score was –0.20 ± 0.58 which was significantly worse ($p = 0.4$) (change toward more positive number is decline). After 6 weeks of placebo washout the score was +0.36 ± 0.62

Comments: It is not entirely clear how many patients' data contributed to the above analysis. The washout of drug effect was similar regardless of treatment sequence, without evidence of a drug carryover effect

Caregiver-rated global impression [No. (%)]	**Donepezil**	**Placebo**	
Improved/total	12/51 (24)	12/53 (23)	Not a significant difference with donepezil vs placebo $p = 0.34$
Worsened/total	14/51 (27)	19/53 (36)	
Explicit verbal memory score mean ± SEM	–0.32 ± 0.28	+0.23 ± 0.29	Differences between the beginning and end of treatment. Positive score indicates improvement
Verbal fluency score, mean ± SEM	–0.71 ± 0.34	–0.27 ± 0.31	

Comments: NYU Stories Test delayed recognition subscale; a 7-point scale where 7 is perfect performance was used to test explicit recall.

Compliance: Dosing compliance was assessed by interview of caregivers and pill counts. Based on pills returned, compliance was estimated as 95.7%

Withdrawals/dropouts	**Group A**		**Group B**	
	Placebo	**Donepezil**	**Donepezil**	**Placebo**
Withdrawal of consent	2	1	1	
Syncope		1		
Recurrent lung cancer		1		
Seizure			1	
Protocol violation			1	
Skin rash				1

Comments: It was judged that the occurrence of syncope (1 patient) and generalised seizure (1 patient) was possibly related to donepezil therapy. An additional patient was diagnosed as having mild pancreatitis at the end of donepezil treatment. None of these complications recurred after discontinuation of donepezil use. Donepezil therapy was otherwise well tolerated.

Adverse effects: Amongst study completers the most common adverse events noted with donepezil therapy were
Nausea 5 of 51 patients (10%)
Diarrhoea 3 of 51 patients (6%)
Agitation 3 of 51 patients (6%)

None of these events were severe or resulted in withdrawal from the study.

Comments: Adverse event monitoring included date of onset and cessation, severity and temporal relation to administration of study medication.

Methodological comments

- Allocation to treatment groups: Treatment group status was assigned by a computerised randomisation schedule generated by a biostatistician. Allocation to treatment group was concealed from all study personnel.
- Blinding: Sites were supplied with sealed opaque individual disclosure forms containing each patient's actual treatment assignment for emergency medical care. Study medications (donepezil or placebo) were packaged in capsules identical in appearance, taste and smell. Cognitive testing was performed by a psychometrician who was masked to the patient's treatment status, adverse event profile, concomitant medication use, pill compliance, caregiver-rated global impression of change, and overall study design.
- Comparability of treatment groups: Not reported but as far as one can tell the treatment groups are comparable.

continued

- Method of data analysis: Outcomes are not reported on an ITT basis, patients who dropped out ($n = 9$) are not included and some were unable to complete testing at a particular visit. In most cases it is clear how many patients' data are being assessed. Results are presented as means with SEM (baseline data means with SDs). Significance testing for repeated measures was performed using analysis of variance, changes in ADAS-cog were also analysed by dividing responses into improvement (negative change) vs no improvement, with significance testing for the within-subject comparison of donepezil and placebo therapy performed using the McNemar test. Differences in caregiver-rated impressions of global change were compared using a 2-sample Wilcoxon test. $p < 0.05$ was required for statistical significance.
- Sample size/power calculation: A target sample size of 60 randomised patients was chosen to provide approximately 80% power to detect a 2.5 point improvement in the ADAS-cog score with donepezil therapy relative to placebo therapy, a magnitude similar to that detected in previous studies of donepezil and tacrine therapy. This calculation assumed a 10% to 20% dropout rate (in the end they had 9 drop out from the starting sample of 60, and only had full data for 48 (80%) participants).
- Attrition/dropout: Attrition/dropout was reported; these were not included in the analysis so the ITT results might be different.

General comments

- Generalisability: Participants were predominantly white and were well educated but the authors believe they were more representative of actual clinical practice that other studies. The authors specifically mention that their study included people who might have been excluded from other trials. Some of the participants were taking psychoactive medication (antidepressants and/or sedatives) and others had significant cardiovascular disease and they responded approximately as well as other participants (data not shown in the paper).
- Outcome measures: Appropriate. There was no specific QoL measure. Unclear whether NYU stories test or verbal fluency test reliable or valid. The authors did attempt to assess compliance by counting pills returned (and estimated compliance to be 95.7%).
- Inter-centre variability: Two centres were included in the study, the Massachusetts General Hospital and the Brigham and Women's Hospital. No variations were reported between these two centres.
- Conflict of interests: None noted.

Quality criteria for Greenberg *et al.*[54]

1. Was the assignment to the treatment groups really random?	Adequate
2. Was the treatment allocation concealed?	Adequate
3. Were the groups similar at baseline in terms of prognostic factors?	Reported
4. Were the eligibility criteria specified?	Adequate
5. Were outcome assessors blinded to the treatment allocation?	Adequate
6. Was the care provider blinded?	Adequate
7. Was the patient blinded?	Adequate
8. Were the point estimates and measure of variability presented for the primary outcome measure?	Adequate
9. Did the analyses include an intention-to-treat analysis?	Inadequate
10. Were withdrawals and dropouts completely described?	Adequate

Reference and design	Intervention	Participants	Outcome measures
Author: Holmes *et al*.[49] Year: 2004 Country: UK Study design: Open label phase followed by RCT Number of centres: 16 Funding: Pfizer/Eisai	Treatment arms: All patients were treated in the open label phase with 5 mg/day donepezil for 6 weeks followed by 10 mg/day donepezil for a further 6 weeks. Patients then randomised to one of the following for a further 6 weeks. Provided there was no marked deterioration (a loss of ≥ 2 points on MMSE compared to baseline) then the randomised treatments continued for a further 6 weeks (1) donepezil 10 mg/day (2) placebo Other interventions used: Concomitant medications were permitted during the study, except other cholinesterase inhibitors	Number of participants: A total of 260 patients were screened of whom 134 entered the study. 96 were randomised Donepezil 10 mg/day = 41 Placebo = 55 Sample attrition/dropout: Of the 41 patients in the donepezil group, 35 (85%) completed and 6 (15%) discontinued (2 had MMSE score drop of 2 or more; 3 due to adverse events; 1 poorly complaint with medication). Of the 55 placebo patients 45 (82%) completed and 10 (18%) discontinued (3 carers removed consent; 1 carer poorly compliant; 6 MMSE dropped by 2 or more) Sample crossovers: none Inclusion/exclusion criteria for study entry: Men and women of any race at least 55 years of age. A diagnosis of probable mild to moderate AD of more than 6 months' duration, as defined by NINCDS-ADRDA criteria. Patients had to have a total NPI score greater than 11 points arising from at least 3 domains of behaviour as assessed by the NPI. Also required the presence of a carer to monitor drug compliance Patients were excluded if they had an MMSE score below 10 or above 27 points, previous exposure to a cholinesterase inhibitor or any clinically relevant disease that may contraindicate their use Characteristics of participants at entry to randomisation: Mean age, years (SE) Gp1 78.6 (1.4); Gp2 78.8 (1.2) Female, *n* (%): Gp1 22 (54); Gp2 37 (67) Mean NPI (SE) Gp1 14.3(1.4); Gp2 15.1 (1.8) NPI median (range) Gp1 13.0 (0-36); Gp2 12.0 (0.62) Mean NPI-D (SE) Gp1 7.5 (0.8); Gp2 7.7 (1.0) Mean MMSE (SE) Gp1 21.1 (0.9); Gp2 20.8 (0.6)	Primary outcomes: Neuropsychiatric Inventory (NPI) Secondary outcomes: Neuropsychiatric Inventory Distress Scale (NPI-D) MMSE Safety Methods of assessing outcomes: The NPI is a carer-based interview that assesses 10 behavioural disturbances in the subject. Severity and frequency of individual behaviours are recorded separately. Severity is rated from 1 (occasional, less than once per week) to 4 (very frequent, daily or continuous). Frequency is rated from 1 (mild) to 3 (severe). The product of severity and frequency ranges from 1 to 12 points for each behaviour assessed with a total score range for all 10 behaviours ranging from 1 to 120 points. The assessment period in all cases was the previous 6 weeks. Weekly diaries, used by the carers, to record neuropsychiatric or other symptoms were encouraged to aid memory The NPI-D is a carer-based interview which assesses the degree of distress caused to the carer by the individual items of behaviour as assessed by NPI. The score for individual items of behaviour ranges from 0 to 5, with a total score range for all 10 behaviours ranging from 0 to 50 points The safety and tolerability of the study medication was assessed continually from baseline to endpoint by monitoring discontinuations from the study and by comparing treatment groups with respect to rates of AEs, concomitant medication use and changes in MMSE Clinic visits occurred at screening, at baseline, and at weeks 6, 12, 18 and 24 Length of follow-up: 24 weeks

continued

Results

Outcomes	**Donepezil 10 mg/day (*n* = 41)**	**Placebo (*n* = 55)**	***p*-Value vs Placebo**
Change in psychometric scores 6 and 12 weeks after randomisation compared with randomisation at Week 12			
NPI (SE)			
6 Weeks	–1.1 (1.5)	5.1 (1.9)	$p = 0.01$
12 Weeks	–2.9 (1.6)	3.3 (2.1)	$p = 0.02$
NPI-D (SE)			
6 Weeks	–0.8 (1.0)	1.9 (1.1)	
12 Weeks	–1.7 (0.7)	1.1 (1.1)	
NPI-D median (range)			
6 Weeks	–1.0 (–16 to 15)	0 (–19 to 28)	$p = 0.03$
12 Weeks	–2.0 (–9 to 10)	1.0 (–20 to 23)	$p = 0.01$
MMSE (SE)			
6 Weeks	0.2 (0.5)	–1.7 (0.4)	$p = 0.005$
12 Weeks	–0.1 (0.6)	–1.8 (0.5)	$p = 0.02$

Comments:
Comparing changes in psychometric scores at randomisation point (week 12) with week 18 there was a decline in NPI total score in the patients allocated donepezil compared with placebo. Likewise, comparing psychometric scores at randomisation point at week 12 with the end of the study at week 24 there was a significant fall in the NPI total score in the patients allocated donepezil compared to those allocated placebo.
Comparing NPI-D scores at randomisation point at week 12 with week 18 there was also a fall in the NPI-D total score in carers of patients allocated donepezil compared to placebo. Likewise there was also a fall in the NPI-D total score in donepezil compared to placebo, when comparing week 12 with the end of study at week 24.
Comparing the MMSE scores at randomisation point at week 12 with week 18 there was an increase in the MMSE total score in patients allocated donepezil compared to those allocated placebo. Likewise, comparing MMSE scores at randomisation point at week 12 with the end of the study at week 24 there was a significant change in the MMSE total score in patients allocated donepezil compared to placebo

	Donepezil 10 mg/day (*n* = 41)	**Placebo (*n* = 55)**	
Mean NPI score ITT-LOCF (estimated from graph)			
Week 18 (6 weeks after randomisation)	13	20	$p < 0.05$
Week 24 (12 weeks after randomisation)	11.5	19	$p < 0.05$
Mean NPI-D score ITT-LOCF (estimated from graph)			
Week 18 (6 weeks after randomisation)	7.0	9.5	$p < 0.05$
Week 24 (12 weeks after randomisation)	6.0	9.0	$p < 0.05$
Mean MMSE ITT-LOCF (estimated from graph)			
Week 18 (6 weeks after randomisation)	21.2	19.1	$p < 0.01$
Week 24 (12 weeks after randomisation)	20.9	19.0	$p < 0.05$

Comments

Adverse effects

Comments: Adverse effects not reported for the randomisation part of the study.

Methodological comments

- Allocation to treatment groups: Patients were randomised using a computer-generated randomisation protocol to placebo or 10 mg/day donepezil on a 3:2 ratio. Randomisation of patients to groups was performed by an independent pharmacist.
- Blinding: The pharmacist provided numbered containers of identical tablets for each patient. All participants were blind to the treatment being offered in the randomisation phase of the study.
- Comparability of treatment groups: Both treatment groups were similar with respect to their demographic characteristics and psychometric test scores at the point of randomisation.

continued

- Method of data analysis: Efficacy analyses were performed to the intent to treat (ITT) population. For the randomised, placebo controlled phase of the study, this was defined as all patients who received at least one dose of study medication at randomisation at week 12, and who provided data at week 12 and at least one post-randomisation efficacy assessment. A randomisation phase analysis comparing the change in NPI and NPI-D total score from the point of randomisation (week 12) to weeks 18 and 24 was undertaken. For the randomised, placebo controlled phase of the study, demographic characteristics and efficacy measure outcomes were measured by unpaired *t*-test for parametric variables and by Mann-Whitney U test for non-parametric variables. All tests were two-tailed and conducted at the 0.05 significance level.
- Sample size/power calculation: Sample size was determined by following a review of the results of phase III clinical trials of donepezil in AD. It was assumed that approximately 1/3 of patients randomised to placebo at 12 weeks would show an appreciable cognitive deterioration at 18 weeks and would therefore be considered a responder to donepezil and would therefore be removed from the study. 121 participants gives 80% power to detect a significant difference of 0.5 SD of the total change in NPI score between groups for the LOCF analysis at 24 weeks. With an anticipated withdrawal rate of 10% during the first 3 months of the study, it was estimated that a minimum of 134 patients would be needed for the baseline population.
- Attrition/dropout: 16 patients (10 patients on placebo and 6 patients on donepezil) withdrew during the randomisation phase of the study. Reasons given. No significant differences in completion rates $p = 0.78$

General comments
- Generalisability: Patients with mild to moderate AD. Lots of dropout between open-label phase and randomised phase, many dropped out due to adverse events.
- Outcome measures: Outcome measures were relevant to the study area and appear to have been measured appropriately.
- Inter-centre variability: Not reported.
- Conflict of interests: The study was supported by an unrestricted project grant from Pfizer/Eisai. Two of the authors had received sponsorship from Pfizer/Eisai to attend educational meetings and as speakers.

Quality criteria for Holmes *et al.*[49]

1. Was the assignment to the treatment groups really random?	Adequate
2. Was the treatment allocation concealed?	Adequate
3. Were the groups similar at baseline in terms of prognostic factors?	Reported
4. Were the eligibility criteria specified?	Adequate
5. Were outcome assessors blinded to the treatment allocation?	Unknown
6. Was the care provider blinded?	Partial
7. Was the patient blinded?	Adequate
8. Were the point estimates and measure of variability presented for the primary outcome measure?	Adequate
9. Did the analyses include an intention-to-treat analysis?	Inadequate
10. Were withdrawals and dropouts completely described?	Adequate

Reference and design	Intervention	Participants	Outcome measures
Author: Homma *et al*.[44] Year: 2000 Country: Japan Study design: RCT, multicentre, double-blind Number of centres: 54 Funding: not reported	Treatment arms: (1) Donepezil 5 mg/day (2) Placebo The trial drugs were administered orally once a day after breakfast for 24 weeks Other interventions used: to avoid gastrointestinal symptoms such as nausea or vomiting during the initial administration phase, the Donepezil patients took a 3-mg tablet for the 1st week Those already undergoing rehabilitation were able to continue as long as there were no changes in the regimen	Number of participants: 268 participants randomised. In the ITT population: donepezil $n = 134$, placebo $n = 129$. Five participants did not undergo efficacy evaluation but unclear which group they were assigned to. There were 228 protocol compatible patients reported (used for baseline and efficacy analysis (donepezil $n = 116$, placebo $n = 112$). This excludes 40 participants but report suggests that 39 were withdrawn (see below) Sample attrition/dropout: 39 patients (15%) were withdrawn: request by family members (10), adverse reactions (8), complications (8), or adventitious diseases (8), did not undergo efficacy evaluation (5) Sample crossovers: none reported Inclusion/exclusion criteria for study entry: outpatients diagnosed with Alzheimer's disease by DSM-IV. Dementia severity of 1 (mild) or 2 (moderate) based on the CDR and cognitive impairment corresponding to an MMSE score of 10–26, and an ADAS-J cog score of ≥ 15 (to exclude those with very mild cognitive impairment). Included if no localised cerebral lesions or multiple infarcts seen by CT and MRI within 6 months before administration of the treatment were considered to be the cause of the dementia Those with Hachinski ischemic score of ≥ 5 points were excluded (to differentiate vascular dementia). Those with neurological signs such as parkinsonism patients with definite symptoms of depression, and patients with old head trauma associated with disturbances of consciousness were also excluded. Also excluded those with visual or hearing impairment, those with aphasia who couldn't undergo the cognitive performance test and patients with no caregivers to provide assistance in outpatient examinations, to assure compliance, and provide reliable information. Patients with serious complications were excluded, including those with peptic ulcers because gastrointestinal bleeding caused by aggravation of peptic ulcers was observed when donepezil was administered in previous clinical trials Concomitant use of choline activators (cholinesterase inhibitors, cholinergic agents), anticholinergics, cerebral vasodilators, activators of cerebral metabolism, psychotropic drugs (major or minor tranquilisers, antidepressants), hypnotics, antiparkinsonism agents, and non steroidal anti-inflammatory drugs prohibited. Initiation of rehabilitation was prohibited	Primary outcomes: ADAS-Jcog (Japan version ADAS-cog); J-CGIC (Japan version CGIC) Secondary outcomes: CDR-SB, MENFIS, CMCS (Caregiver-rated Modified Crichton Scale – a modified Crichton Geriatric Rating Scale (CGRS) for activities of daily living, adverse events, compliance (also laboratory tests but not data extracted as per protocol) Methods of assessing outcomes (see results section for scoring): ADAS-Jcog reliability and validity confirmed. Implemented by the same clinical psychologist or speech/hearing therapist during the study (training given to ensure a uniform evaluation) J-CGIC was adapted for the study, measured by the study investigator. Paper discusses that the CGIC has had low reliability in the past and that the CIBIC-plus has been developed; however, the current conditions in Japan at the time of the trial didn't allow for a clinician other than the investigator to make the evaluations (necessary for the CIBIC-plus) and therefore a new CGIC was prepared CDR measures severity of dementia and the Japanese version has been confirmed to have satisfactory interrater reliability. MENFIS is a modification of the GBS scale. Reliability and validity are confirmed. CDR-SB, MENFIS, were measured by the study investigators and CMCS (with the CGRS) by caregiver diary, some amendments to the functions assessed (see results) and reliability and validity not assessed

continued

Reference and design	Intervention	Participants	Outcome measures

Characteristics of participants in protocol compatible patients (mean ± SD (range) unless stated):

	Donepezil 5 mg $n = 116$	Placebo $n = 112$	p-Value
M/F	37/79 (32%/68%)	38/74 (34%/66%)	0.853
Age, years	70.1 ± 7.6 (52–83)	69.4 ± 8.8 (48–90)	0.521
Weight, kg	51.3 ± 8.4 (33–70)	50.0 ± 9.3 (29–73)	0.316
Severity,	*n* (%)	*n* (%)	
CDR-1	79 (68%)	69 (62%)	
CDR-2	37 (32%)	43 (38%)	0.305
MMSE†	17.8 ± 3.9 (10–26)	16.6 ± 3.9 (10–26)	0.035*
ADAS-J†	22.91 ± 8.49 (15.0–56.7)	26.90 ± 9.84 (15.0–60.0)	0.001*

*$p < 0.15$, †ITT mean ± SE reported below

Compliance: recovery of residual drug from the caregiver on hospital visits every 4 weeks and determining actual number of tablets taken using caregivers' diaries. Compliance rates for each administration and the whole administration period were calculated (only those with ≥2/3 in both were included in analysis, as a rule)

Patients were examined every 4 weeks
Length of follow-up: 24 weeks

Results (Where 'ITT' and PC results presented in report, only 'ITT' data extracted)

Outcomes ('ITT' population)	Donepezil ($n = 134$)	Placebo ($n = 129$)	Drug-placebo difference
ADAS-J cog mean ± SE baseline (range), *n*	23.00 ± 0.73 (8.0–56.7), $n = 133$	26.73 ± 0.88 (14.0–60.0), $n = 126$	
Mean change ± SE from baseline	–2.43 ± 0.45, $n = 126$	0.11 ± 0.49, $n = 113$	–2.54, $p = 0.001$*
Mean change with time (PC population) estimated from figure			
4 weeks	–1.2	–0.5	
8 weeks	–2.4	–1.1	
12 weeks	–3.0	–0.9	$p = 0.008$
16 weeks	–3.3	–0.8	$p = 0.001$*
20 weeks	–3.4	0.1	$p = 0.001$*
24 weeks	–3.1	0.1	$p = 0.003$
Endpoint see above			
ADAS-Jcog scoring method not reported; change score minus = improvement; *paper reports $p = 0.000$ but not statistically plausible			
J-CGIC			
Improvement rates (% slightly improved or better)	52%	22%	
Aggravation rates (% slightly aggravated or worse)	17%	43%	
N (%) change ('ITT')	$n = 134$ (1 unassessable)	$n = 129$ (1 unassessable)	
Markedly improved	1 (1)	2 (2)	
Improved	21 (16)	13 (10)	
Slightly improved	42 (31)	10 (8)	
No change	44 (33)	49 (38)	$p = 0.001$*
Slightly aggravated	19 (14)	22 (17)	
Aggravated	6 (4)	25 (19)	
Markedly aggravated	0 (0)	1 (1)	
Difference in improvement ratio, 95% CI (proportion 'improved' or better to 'total')	ITT 4.8% (–3.6, 13.2%)		

continued

Difference in aggravation ratio, 95% CI [proportion ('slightly aggravated' or worse + unassessable) to 'total']	ITT 23.2% (12.4, 34.1%)

J-CGIC based on a Likert-type assessment and consists of a global assessment of changes in the patient's clinical symptoms subjectively by clinicians into seven grades (1 = markedly improved; 2 = improved; 3 = slightly improved; 4 = no change; 5 slightly aggravated; 6 = aggravated; 7 = markedly aggravated); *reports that the efficacy in the donepezil group was confirmed with $p = 0.000$ but not statistically plausible.

CDR-SB (PC population)	***n* = 116**	***n* = 112**	**Drug-placebo difference**
Mean ± SE baseline (range)	7.17 ± 0.21 (4.0–12.0)	7.55 ± 0.22 (3.5–12.0)	
Mean change ± SE from baseline	–0.10 ± 0.12	0.75 ± 0.15	–0.85, $p = 0.001$*
Mean change with time estimated from figure			
4 weeks	0	0	
8 weeks	–0.1	0.05	
12 weeks	–0.15	0.2	$p = 0.018$
16 weeks	–0.2	0.4	$p = 0.001$*
20 weeks	–0.25	0.5	$p = 0.001$*
24 weeks	–0.15	0.7	$p = 0.001$
Endpoint see above			

CDR-SB scores range from 0 to 18 (not stated which is more severe, assume lowest as on change scores minus = improvement).

MENFIS (PC population)	***n* = 116**	***n* = 112**	**Drug-placebo difference**
Mean ± SE baseline (range)	27.28 ± 0.85 (9–54)	30.13 ± 0.86 (11–49)	
Mean change ± SE from baseline	–0.72 ± 0.53	1.84 ± 0.69	–2.56, $p = 0.004$
Mean change with time estimated from figure			
4 weeks	–0.7	–0.4	
8 weeks	–1.2	–0.3	
12 weeks	–1.25	0.4	$p = 0.021$
16 weeks	–1.2	1.3	0 = 0.011
20 weeks	–1.3	1.7	$p = 0.002$
24 weeks	–1.1	1.75	$p = 0.008$
endpoint see above			

MENFIS is a modification of the GBS scale, aimed at evaluating the core symptoms of dementia syndromes including cognitive, motivational and emotional functions. The scoring of each function is 0 to 42 for cognitive function, 0 to 18 for motivational function, and 0 to 18 for emotional function; therefore the range of the total score is from 0 to 78. The higher the score the greater the degree of functional deficit (minus change = improvement).

CMCS (PC population)	***n*s different for each**	***n*s different for each**	**Drug-placebo difference**
Mean ± SE baseline (range), *n*	18.63 ± 0.77 (0–40), *n* = 109	19.33 ± 0.76 (3–38), *n* = 108	
Mean change ± SE from baseline, *n*	1.03 ± 0.66, *n* = 103	3.45 ± 0.71, *n* = 99	–2.42, $p = 0.01$
Mean change with time estimated from figure			
4 weeks	0.01	–0.03	
8 weeks	0	0.02	
12 weeks	–0.05	1.5	$p = 0.006$
16 weeks	0	2.6	$p = 0.003$
20 weeks	0.05	3.4	$p = 0.008$
24 weeks	0.07	3.5	$p = 0.009$
Endpoint see above			

continued

CMCS: A minus change score = improvement. Aggravation occurred in both groups but the degree of aggravation was less in the donepezil group than the placebo group. This included the CGRS which consists of 11 items; 4 related to ADL: mobility, dressing, feeding, continence, 3 related to communication: orientation, conversation, cooperation, and 4 related to psychiatric symptoms: restlessness, sleep, objective mood, subjective mood. However the ADL items were evaluated as "normal" in ≥80% of the patients from before administrations from the results of a late phase II trial in Japan, and were therefore deemed not suitable for evaluation of drug efficacy in patients with mild to moderately severe AD. Also objective mood and subjective mood were deemed to be essentially mental state assessments intended for specialists, and these were removed and replaced with 'work and social activities' and 'leisure'. As a result a total of 7 items were evaluated in 8 grades, the scoring range was therefore 0 to 56, and the higher the score the greater the degree of deficit

Adverse effects	***n* = 136**		***n* = 131 (1 excluded)**		**Treatment vs placebo**	
Patients stopping treatment due to adverse events	2 (1%)		6 (5%)			
Drug related incidence	10% (14/136)*		8% (10/131)		$p = 0.587$	
Those with ≥3 incidences	**adverse events**	**adverse reaction†**	**adverse event**	**adverse reaction†**	**adverse events**	**adverse reaction†**
Total patients showing AE	54 (40%)	14 (10%)	33 (25%)	10 (8%)	0.016	0.587
Total number of events	79	16	53	15		
Gastrointestinal disorders:						
diarrhoea	5 (4%)	3 (2%)	4 (3%)	1 (1%)	1.000	0.651
nausea	6 (4%)	3 (2%)	1 (1%)	1 (1%)	0.133	0.651
abdominal pain	2 (1%)	1 (1%)	3 (2%)	1 (1%)	0.965	1.000
vomiting	2 (1%)	1 (1%)	2 (2%)	2 (2%)	1.000	0.972
anorexia	2 (1%)	1 (1%)	2 (2%)	2 (2%)	1.000	0.972
constipation	2 (1%)	1 (1%)	1 (1%)	0	1.000	0.517
Mental and neurological disorder:						
restlessness	0	0	3 (2%)	2 (2%)	0.233	0.480
Central or peripheral nerve disorder:						
headache	4 (3%)	0	1 (1%)	0	0.394	1.000
Others:						
cold syndrome	10 (7%)	0	2 (2%)	0	0.040	1.000
inflammation upper airway	3 (2%)	0	2 (2%)	0	1.000	1.000
fever	3 (2%)	0	2 (2%)	0	1.000	1.000
fracture	1 (1%)	0	3 (2%)	0	0.594	1.000
eczema	3 (2%)	0	0	0	0.261	1.000

Comments: adverse events listed using the terminology in 'Adverse Drug Reaction Terminology' (1996 ed).

*The main drug-related adverse events in the donepezil group were gastrointestinal symptoms (see above) All of these were mild or moderate, all disappeared when the trial drug was withdrawn or temporarily discontinued.

†adverse events for which the causal relationship to the investigational drug could not be denied (other than 'not related') were defined as adverse reactions. Excluding cold syndrome, incidence by symptoms showed no intergroup difference

Compliance	98% reached the specified compliance rate for the efficacy analysis

Methodological comments

- Allocation to treatment groups: States participants were randomly allocated but no further details. No details of concealment of allocation.
- Blinding: Describes the trial as double-blind but no further details. No details of blinding of outcome assessors.
- Comparability of treatment groups: There were differences between groups in the scores obtained by each scale (MMSE, ADAS-Jcog, MENFIS) before treatment (reports that an analysis showed that this had no effect on the interpretation of the trial results).

continued

- Method of data analysis: Reports that analysis of patient demographics and efficacy performed on a protocol compatible (PC) population and an ITT population. The ITT population was defined as a population with no problems in obtaining consent to participate in the trial, which was diagnosed as having dementia of the AD type, was examined at least once after administration of the trial drug, and underwent some kind of efficacy evaluation (therefore not by definition ITT as some patients may have dropped out before the first efficacy evaluation). The PC population was defined as a population which was handled in accordance with the provisions in the clinical trial protocol. Reports that it was confirmed that interpretation of the results did not differ between the two populations. Primary outcomes were assessed by the U test, secondary efficacy analysis by the U test for score differences between baseline and each measurement time. Intergroup differences tested. Adverse events assessed using Fisher's exact test. The level of significance was 5% for all cases (except baselines 15%), and the *p* was rounded to 3 decimal places.
- Sample size/power calculation: Not described.
- Attrition/dropout: numbers and reasons given for withdrawals (although appears to be one not accounted for) but no clear numbers provided for many that were not included in the assessments of effectiveness (lost to follow-up).

General comments
- Generalisability: Assume Japanese population, those with mild to moderate AD, excluded those with likely vascular dementia.
- Outcome measures: Adapted measures for Japanese population. primary outcome measures reliable and valid, unclear with secondary measures.
- Inter-centre variability: Not reported.
- Conflict of interests: Funding information not given.

Quality criteria for Homma *et al.*[44]

1. Was the assignment to the treatment groups really random?	Unknown
2. Was the treatment allocation concealed?	Unknown
3. Were the groups similar at baseline in terms of prognostic factors?	Reported
4. Were the eligibility criteria specified?	Adequate
5. Were outcome assessors blinded to the treatment allocation?	Unknown
6. Was the care provider blinded?	Partial
7. Was the patient blinded?	Partial
8. Were the point estimates and measure of variability presented for the primary outcome measure?	Adequate
9. Did the analyses include an intention-to-treat analysis?	Inadequate
10. Were withdrawals and dropouts completely described?	Partial

Reference and design	Intervention	Participants	Outcome measures
Author: Krishnan *et al*.[45] Year: 2003 Country: USA Study design: RCT, double-blind multicentre Number of centres: 3 Funding: Pharmaceutical companies: Eisai Inc and Pfizer Inc.	Treatment arms: (1) Donepezil, once daily each evening. For the first 28 days 5 mg/day and then 10 mg/day thereafter (2) Placebo, once daily each evening Daily doses consisted of two identical tablets so as not to reveal the titration scheme: two placebo tablets throughout the study for the patients receiving placebo; one placebo and one 5-mg donepezil tablet or two donepezil tablets for the donepezil group Other interventions used: concomitant pyschotrophic drugs were not allowed	Number of participants: 67: donepezil (group 1) $n = 34$, placebo (group 2) $n = 33$ Sample attrition/dropout: Donepezil (6 discontinued: 2 request of patient or investigator; 2 protocol violation; 2 other). Placebo (10 discontinued: 1 adverse events; 3 request of patient or investigator; 1 protocol violation; 5 other). Other = unacceptable baseline. Those withdrawing were requested to return for a final evaluation including psychometric assessment, adverse events, and an MRI/H-MRS scan Sample crossovers: none Inclusion/exclusion criteria for study entry: women (at least 2 years post menopausal or surgically sterile) and men ≥ 50 years with a diagnosis of probable, mild-to-moderate, uncomplicated Alzheimer's disease according to DSM-IV and NINCDS criteria. A Clinical Dementia rating of 1 (mild) or 2 (moderate), an MMSE of 10–26, and a Hachinski score of ≤ 4 required at screening and baseline. In addition, patients had to be in generally good health, ambulatory, and with sufficient hearing and vision for compliance with testing procedures. Only those able to undergo MRI examination (i.e. those without a pacemaker or other metal items within the body; those who were not claustrophobic) were enrolled. Patients with a primary diagnosis of psychiatric disorders other than AD, cerebrovascular disease, or any unstable medical conditions were excluded Characteristics of participants (mean ± SD unless stated: Age (years): group 1: 74.4 ± 7.0; group 2: 72.4 ± 10.1 Female, *n* (%): group 1: 25 (74); group 2: 23 (70) White race, *n* (%): group 1: 34 (100); group 2: 30 (91) Black race, *n* (%): group 1: 0; group 2: 3 (9) CDR 1 (mild), *n* (%): group 1: 26 (76); group 2: 25 (76) CDR 2 (moderate), *n* (%): group 1: 7 (21); group 2: 8 (24) CDR 3 (severe), *n* (%): group 1: 1 (3), group 2: 0 MMSE: group 1: 19.5 ± 4.8 (range 10–26); group 2: 19.0 ± 4.6 ($n = 33/34$) (range 10–25) ADAS-cog: group 1: 26.51 ± 12.13; group 2: 26.44 ± 12.29 ($n = 33/34$)	Primary outcomes: brain *n*-acetylaspartate concentrations (not data extracted as per protocol) Secondary outcomes: ADAS-cog and adverse events (also hippocampal volume, brain myo-inositol concentrations, safety by physical examination, clinical laboratory tests but not data extracted as per protocol) Methods of assessing outcomes: ADAS-cog assessed using the 11-item scale which tests cognition such as memory, language and praxis functions (see below for scoring). Assessment was conducted by a trained clinical staff member Patients were required to return at 6 week intervals (± 3 days) for the following evaluations: routine physical exam, lab assessments, ECG, MRI/H-MRS scan, psychometric assessment, medication, compliance check, adverse event monitoring Length of follow-up: 24 weeks (followed by 6 week single-blind placebo washout period)

continued

Results			
Cognition	**Donepezil *n* = 34**	**Placebo *n* = 32**	***p*-Value between groups**
ADAS-cog (change from baseline, estimated from figure)			
6 weeks	–3.1 (*n* = 34)	0.4 (*n* = 32)	$p < 0.003$
12 weeks	–2.1 (*n* = 31)	1.4 (*n* = 30)	$p < 0.007$
18 weeks	–1.7 (*n* = 30)	1.8 (*n* = 29)	$p < 0.04$
24 weeks	–0.6 (*n* = 28)	3.3 (*n* = 28)	$p < 0.02$
Endpoint (estimated)	0.2 (*n* = 34)	3.2 (*n* = 32)	$p < 0.04$
Comments: ADAS-cog-11 scores range from 0–70 with increasing score denoting worsening (positive change = clinical decline).			
Adverse effects (incidence)	94%	85%	ns
Comments: reports no statistically significant differences between the two groups in the incidence of specific events, but no further details. No patients in the donepezil and 1 patient in the placebo group discontinued due to adverse events			

Methodological comments

- Allocation to treatment groups: Patients were randomly assigned by means of a computerised randomisation schedule; not clear whether allocation was concealed.
- Blinding: Described as double-blind, daily doses of identical tablets were used. Outcome assessors not blinded.
- Comparability of treatment groups: Groups reported to be comparable on age, gender, severity of symptoms, and race.
- Method of data analysis: Observed case analyses and endpoint analyses were used for the ADAS-cog (endpoint was the last post-baseline valued carried forward for patients missing a week 24 assessment). Differences between groups for the mean post-baseline changes in scores on the ADAS-cog were compared by ANCOVA. Incidences of adverse events were analysed using Fisher's exact test.
- Sample size/power calculation: This is a preliminary study and was designed for the primary outcome of brain *N*-acetylaspartate concentrations. No power calculation reported.
- Attrition/dropout: Reported and reasons given. Evident that some unavailable for assessment but reasons not reported.

General comments

- Generalisability: Those over 50 years with mild-moderate AD. Note that one patient in the donepezil group had a CDR score = severe dementia.
- Outcome measures: Appropriate.
- Inter-centre variability: Not reported.
- Conflict of interests: Funded by Eisai Inc, Teaneck NJ, and Pfizer Inc. NY. Five authors have received grants and honoraria from Eisai Inc, and Pfizer Inc. Two authors are employees of Eisai Inc. One was an employee of Eisai Inc. when the study was undertaken.

Quality criteria for Krishnan *et al.*[45]

1. Was the assignment to the treatment groups really random?	Adequate
2. Was the treatment allocation concealed?	Unknown
3. Were the groups similar at baseline in terms of prognostic factors?	Reported
4. Were the eligibility criteria specified?	Adequate
5. Were outcome assessors blinded to the treatment allocation?	Unknown
6. Was the care provider blinded?	Adequate
7. Was the patient blinded?	Adequate
8. Were the point estimates and measure of variability presented for the primary outcome measure?	Inadequate
9. Did the analyses include an intention-to-treat analysis?	Inadequate
10. Were withdrawals and dropouts completely described?	Partial

Reference and design	Intervention	Participants	Outcome measures
Author: Mohs *et al.*[46] Year: 2001 Country: USA Study design: RCT, double-blind multicentre Number of centres: 31 Funding: Jointly sponsored by Eisai Inc. and Pfizer Inc.	Treatment arms: (1) 5 mg/day donepezil (28 days); then 10 mg/day thereafter, up to 54 weeks (2) Placebo Other interventions used: Use of vitamin E, Ginko biloba, non-steroidal anti-inflammatory drugs (NSAID) and oestrogens was permitted. (Concomitant use of anticholinergenics, cholinomimetics, tricyclic antidepressants, antiparkinson agents, and neuroleptics was not permitted)	Number of participants: total group 431. group 1) $n = 214$; group 2) $n = 217$; however 16 patients excluded from ITT population; 7 in donepezil and 9 placebo (baseline assessments missing or no post-baseline assessment) Sample attrition/dropout: (1) discontinued prematurely $n = 60$ (adverse events: 23 (10.7%) (includes 3 who died), request of subject/investigator: 14 (6.5%), medication non-compliance: 3 (1.4%), protocol violation: 3 (1.4%), entered nursing facility: 2 (0.9%), other: 15 (7.0%)); completed 54 weeks on assigned treatment $n = 68$ (2) discontinued prematurely $n = 56$ [adverse events: 16 (7.4%) (includes 4 who died), request of subject/investigator: 11 (5.1%), medication non-compliance: 1 (0.5%), protocol violation: 7 (3.2%), entered nursing facility: 2 (0.9%), other:19 (8.8%)]; completed 54 weeks on assigned treatment $n = 43$ 'Other' includes lost to follow-up; prescribed donepezil; withdrawal of consent; decision by physician; minimum improvement per caregiver; off study drug too long in error; removed for meeting criteria for functional decline in error; criteria for functional decline missed in error	Primary outcomes: Alzheimer's Disease Functional Assessment and Change Scale (adapted for the study) (ADFACS) assessing functional impairment, based on 6 basic ADL items (toileting, feeding, dressing, personal hygiene and grooming, bathing, walking) and 10 instrumental items (use of telephone, household tasks, using household appliances, managing money, shopping, food preparation, ability to get around inside and outside home, hobbies and leisure activities, handling personal mail, grasp of situations or explanations); Clinical Dementia Rating (CDR) scale assessing the severity of dementia; Mini-Mental State Examination (MMSE) evaluating cognitive state. Clinically evident decline in function Secondary outcomes: Safety monitored by evaluating the incidence of adverse events through scheduled questioning and spontaneous reporting. In all cases, judgement of the relationship of study treatment to an adverse event was made by the investigator under double-blind conditions. Also included physical examination, vital signs, clinical laboratory tests, and ECG. Data not extracted as per protocol

continued

Reference and design	Intervention	Participants	Outcome measures
		Most patients who were discontinued from the study did so because they met the criteria for clinically evident functional decline Sample crossovers: none Inclusion/exclusion criteria for study entry: All patients were required to have a diagnosis of probable AD (DSM-IV and NINCDS) as well as an MMSE score of 12–20 inclusive, CDR score of 1 (mild) or 2 (moderate), and modified Hachinski ischaemia scores of ≤ 4 at both screening and baseline. (A protocol amendment allowed patients to enrol with MMSE scores of 21 at baseline if their scores at screening were 20 (20/21 criteria). Subjects also required to perform 8 of 10 instrumental ADL (each score ≤ 2) and 5 of 6 basic ADL (each score ≤ 2) on the ADFACS at both screening and baseline. AD functional assessment and change score Patients were excluded if evidence of other neurologic or psychiatric disorders (i.e.: stroke, Parkinson's disease, schizophrenia), dementia complicated by other organic disease, delirium (DSM-IV) 290.30 or 290.11), depression (DSM-IV 290.21 or 290.13), or AD with significant delusions (DSM-IV categories of 290.20 or 290.12). Additional exclusion included history of alcoholism or drug misuse, hypersensitivity to cholinesterase inhibitors, or use of any investigational drug or tacrine within 1 month of screening. Patients were also excluded if they were without a reliable caregiver Characteristics of participants: (see below)	Patients attended the clinic for efficacy and safety evaluations at screening (within 30 days prior to drug administration), at baseline, and at 6 week intervals (42 ± 3 days) throughout the study, up to 54 weeks (day 379 ± 3 days, final visit), or at an unscheduled termination visit Methods of assessing outcomes: All assessments were carried out by investigators who were blind to the patient's treatment. The MMSE, evaluating cognitive state, was administered to the patient by a trained clinician The CDR ratings for cognition (memory, orientation, judgement and problem solving) and function (community affairs, home and hobbies, personal care) were assessed by consensus of the patient's assessment team, including the caregiver Length of follow-up: 54 weeks

continued

Baseline characteristics

Patient characteristics	Donepezil 10 mg/day (n = 214)	Placebo (n = 217)
Age; mean (SE), range	75.4 (0.6), 50–91	75.3 (0.6), 49–94
Gender, *n* (%)	Female: 131 (61.2); Male: 83 (38.8)	Female: 140 (64.5); Male: 77 (35.5)
Race, *n* (%)*	White: 203 (94.9); Black: 2 (0.9); other: 9 (4.2)	White: 194 (89.4); Black: 10 (4.6); 13 (6.0)
Weight kg, mean (SE), range	67.0 (1.0), 35.0–121.0	66.3 (1.0), 38.2–118.0
Height, cm mean (SE), range	164.3 (0.7), 125.0–188.6	163.8 (0.7), 129.5–189.2
Baseline CDR rating, n (%)	1.0: 174 (81.3); 2.0: 40 (18.7)	1.0: 174 (80.2); 2.0: 43 (19.8)
Baseline CDR-SB, mean (SE)†	6.81 (0.14)	6.78 (0.15)
Baseline MMSE§**, mean (SE), range	17.1 (0.2), 11–30	17.1 (0.2), 11–23

Notes: *p = 0.04 Between treatment groups; †Data for intention to treat population; §Investigator-derived scores were used in the analysis, but one donepezil patient who was scored as 30 actually had a score of 15.
** 8 donepezil and 10 placebo patients deviated from the original MMSE criteria. 13/18 were within 2 points, usually with higher scores.

Results

Clinically evident functional decline	Donepezil 10 mg/day (n = 207)	Placebo (n = 208)
Number of patients meeting criteria for clinically evident functional decline	41%, n = 84 of the ITT population	56%, n = 116 of ITT population
Median time to clinically evident functional decline (days; 95% CI)*	357; lower limit of the 95% CI = 280 days	208; 95% CI: 165, 252 days
Probability of survival with no clinically evident functional decline at 48 weeks; (95% CI)**	51% (43%, 58%)	35% (27%, 42%)
Number of patients meeting criteria for clinically evident functional decline as result of decline in CDR	17	17

Comments:
Criteria for clinically evident decline in function (any one of the following):
(1) A clinically evident decline in ability to perform one or more basic ADL (ADFACS) present at baseline. A clinically evident decline was defined as a decline of at least one point, except that a decline from 0 (no impairment) to 1 (mild impairment) was not considered clinically significant.
(2) A clinically evident decline in the ability to perform 20% or more of the instrumental ADL (ADFACS) present at baseline. A decline from 0 (no impairment) to 1 (mild impairment) was not considered clinically significant but other declines of one or more points were.
(3) An increase in global CDR score of 1 point or more compared with baseline (e.g.: from a score of 1 to 2 to 3, or from 2 to 3).
*Donepezil patients maintained their function 72% longer than those on placebo.
** The hazard ratio for reaching endpoint (donepezil/placebo) was 0.62. Thus, patients treated with donepezil were 38% less likely to decline over a 1-year period.

	Donepezil (1)	Placebo (2)	p-Value
Adjusted mean change from baseline to week 54 and endpoint in ADFACS total score for patients remaining on assigned treatment; estimated from figure (patients evaluated; *n*)*	Week 0: 0 (183); Week 6: 0.45 (181); Week 12: 0.4 (150); Week 18: –0.1(123); Week 24: –0.3 (97); Week 30: –0.2 (85); Week 36: –0.35 (74); Week 42: –0.15 (69); Week 48: –0.5 (61); Week 54: 0.3 (61); Endpoint: 2.4 (181)	Week 0: 0 (197); Week 6: 0.6 (197); Week 12: 1.3 (171); Week 18: 0.7 (125); Week 24: 0.9 (94); Week 30: 0.45 (70); Week 36: 0.65 (60); Week 42: 1.15 (54); Week 48: 0.1 (47); Week 54: 0.15 (41); Endpoint: 3.85 (196)	$p < 0.01$ (1 versus 2) $p < 0.01$ (1 versus 2) $p < 0.05$ (1 versus 2) $p < 0.05$ $p < 0.001$ (1 versus 2)

continued

	Donepezil (1)	Placebo (2)	*p*-Value
Adjusted mean change from baseline to week 54 and endpoint in MMSE for patients remaining on assigned treatment; estimated from figure (patients evaluated; *n*)**	Week 0: 0 (207); Week 6: 1.1 (205); Week 12: 1.45 (171); Week 18: 1.75 (138); Week 24: 1.8 (111); Week 30: 1.5 (99); Week 36: 1.95 (84); Week 42: 1.05 (79); Week 48: 1.8 (70); Week 54: 1.3 (68); Endpoint: 0.6 (207)	Week 0: 0 (208); Week 6: –0.1 (208); Week 12: –0.15 (178); Week 18: 0.6 (127); Week 24: 0.45 (96); Week 30: 0.4 (73); Week 36: 0.95 (62); Week 42: –0.3 (56); Week 48: 1.1 (49); Week 54: 0.55 (43); Endpoint: –0.6 (208)	 $p < 0.001$ (1 versus 2) $p < 0.001$ (1 versus 2) $p < 0.01$ (1 versus 2) $p < 0.01$ (1 versus 2) $p < 0.001$ (1 versus 2)

Comments:
Each of the basic ADL items is scored on a scale of 0 (no impairment) to 4 (very severe impairment), giving a basic ADL total score range of 0 to 24. Each of the instrumental ADL items is scored on a scale of 0 (no impairment) to 3 (severe impairment), giving an instrumental ADL total score range of 0 to 30. The overall 16-item ADFACS total score has a range of 0 to 54 (best to worse).
CDR scale provides a global rating of the severity of dementia on a 5-point scale ranging from 0 (normal, no impairment) to 3 (severe impairment).
CDR-SB calculated as the sum of the ratings for each of the six CDR domains (boxes).
At each visit, investigators determined whether predefined criteria for clinically evident decline in functional status had been met. Patients who met the endpoint criteria were discontinued per protocol.
*The adjusted mean changes favoured donepezil. At study endpoint (last recorded visit for all patients), the differences in mean change from baseline for donepezil patients differed from placebo for both instrumental ADL ($p = 0.001$) and basic ADL ($p = 0.007$). Analysis of mean change from baseline to study endpoint for the 10 instrumental ADL items on the ADFACS showed less impairment of function with donepezil compared with placebo. With regards to the individual basic ADL items, the mean change from baseline to endpoint showed less decline with donepezil compared with placebo for 5/6 items, reaching significance in favour of donepezil for feeding and dressing.
**CDR-SB total scores were consistently lower and MMSE consistently higher in the donepezil group compared with placebo at all scheduled visits for patients remaining on treatment. Significant differences in favour on the CDR-SB at weeks 6, 18, 24, 36 and 42

	Donepezil	Placebo	
Adverse effects experienced by at least 5% of all patients taking donepezil	Overall *n* (%); severe, *n*; related,* *n*	Overall *n* (%); severe, *n*; related, *n*	*p*-Value versus placebo
Accidental injury	12 (6); 1; 0	6 (3); 2; 1	0.16
Asthenia	14 (7); 1; 8	8 (4); 0; 4	0.20
Headache	20 (9); 0; 10	7 (3); 1; 4	0.01
Anorexia	12 (6); 0; 9	4 (2); 0; 0	<0.05
Diarrhoea	37 (17); 1; 25	11 (5); 0; 9	<0.001
Dyspepsia	12 (6); 0; 0	3 (1); 0; 0	<0.05
Nausea	19 (9); 0; 14	8 (4); 0; 6	<0.05
Weight loss	13 (6); 0; 11	9 (4); 0; 7	0.39
Agitation	28 (13); 1; 13	21 (10); 0; 10	0.29
Insomnia	16 (8); 0; 7	7 (3); 0; 3	0.06
Rhinitis	25 (12); 1; 5	14 (7); 0; 1	0.07
Abrasion	16 (8); 0; 1	7 (3); 0; 0	0.06
Urinary tract infection	28 (13); 0; 1	14 (7); 0; 0	<0.05
Number of deaths	3 (unrelated to study medication)	4 (3 of which unrelated to study medication)	

Comments:
* Includes events judged by the investigator to be either possibly or definitely related to test drug. The majority of adverse events were mild to moderate in intensity and unrelated or possibly related to study medication.
The difference in rates of premature discontinuations for adverse events, were small and consistent throughout the trial.
Compliance had to be ≥ 80% for a patient's visit to be considered evaluable. Patients could not continue if they had two or more consecutive non-evaluable visits.
26 donepezil patients (12.1%) experienced 43 serious adverse events and 19 patients (8.8%) treated with placebo experienced 26 serious adverse events. Of these serious AEs, 36/43 in the donepezil group were judged by the investigator not to be related to study medication. The remaining 7 thought to be related to medication were: syncope (3 patients; 2 rated as severe; all hospitalised), and breast neoplasm, agitation, anxiety and apnea (one patient each).

continued

Methodological comments

- Allocation to treatment groups: Patients were randomised in blocks of four to receive once daily doses of donepezil or placebo. Randomisation method not reported. No details of concealment allocation.
- Blinding: Study described as double-blind. All assessments were carried out by investigators who were blind to the patient's treatment. The blind was broken for one patient who had an adverse event, but this did not compromise allocation concealment.
- Comparability of treatment groups: There were no significant treatment differences for demographic characteristics except race. The difference between treatment groups was due primarily to 10 of the 12 African-American subjects being randomised by chance to the placebo group. Questionable therefore, as to whether randomisation was indeed correct.
- Method of data analysis: The efficacy analyses were based on the ITT patient population, defined as all subjects randomised to treatment who received at least one dose of double-blind medication and who had baseline and at least one post-baseline assessment of efficacy. The Kaplan-Meir method was used to obtain survival time estimates. Log-rank and Wilcoxon tests for difference in the survival distributions (or proportional hazard model to estimate the hazard ratio of functional decline).
- Sample size/power calculation: The sample size of 200 per treatment group was determined based on the estimated 1-year values for significant functional decline of 55% for the placebo group and 38% for the 10mg/day donepezil group, extrapolated from a previous 6-month study. For this sample size, the power was calculated to be 0.91.
- Attrition/dropout: 1) 68 completed 54 weeks on assigned treatment; 2) 43 completed 54 weeks on assigned treatment.

General comments

- Generalisability: Mild to moderate AD.
- Outcome measures: Consisted of the ADFACS, the CDR scale, and the MMSE. The MMSE was adapted, therefore its reliability and validity are questionable.
- Inter-centre variability: None reported; unclear how many patients from each centre.
- Conflict of interests: Three of the authors had received compensation from Eisai Inc and Pfizer Inc as consultants and lecturers. Two authors were current employees of Eisai Inc., Teaneck, NJ, and one author was an employee of Eisai Inc during the conduct of the study.

Quality criteria for Mohs *et al.*[46]

1. Was the assignment to the treatment groups really random?	Unknown
2. Was the treatment allocation concealed?	Unknown
3. Were the groups similar at baseline in terms of prognostic factors?	Reported
4. Were the eligibility criteria specified?	Adequate
5. Were outcome assessors blinded to the treatment allocation?	Partial
6. Was the care provider blinded?	Partial
7. Was the patient blinded?	Partial
8. Were the point estimates and measure of variability presented for the primary outcome measure?	Partial
9. Did the analyses include an intention-to-treat analysis?	Inadequate
10. Were withdrawals and dropouts completely described?	Adequate

Reference and design	Intervention	Participants	Outcome measures
Author: Nunez *et al*.[48] (poster) (AWARE study) Year: 2003 Country: Multinational Study design: RCT (following an open label phase) Number of centres: Not specified. However, this was a multinational trial with patients recruited from Belgium, Denmark, Germany, Greece, Hungary, Poland, the Netherlands and the USA Funding: Not reported	Treatment arms: Following an open label trial of 12–24 weeks donepezil, patients showing 'no apparent clinical benefit' were randomised into one of the following: (1) 10 mg/day donepezil (2) Placebo Other interventions used: none reported	Number of participants: 817 were enrolled into initial study (5 did not take study drug and 193 withdrew). 193/619 (31.2%) patients reported as having 'no apparent clinical benefit'. 15 patients were incorrectly randomised from the 'observed clinical benefit group' and 6 patients declined to be randomised from the 'no apparent clinical benefit group'. 202 patients were therefore randomised. Donepezil 10 mg/day = 99 Placebo = 103 Sample attrition/dropout: not reported Sample crossovers: none Inclusion/exclusion criteria for study entry: Patients with mild to moderate (MMSE 10–26), possible or probable AD (DSM IV and NINCDS-ADRDA criteria) The current use (within 30 days or 5 half lives) of any investigational or approved drugs for AD was an exclusion criteria. Characteristics of participants: Male *n* (%) Gp1 40 (40.4); Gp2 38 (36.9) Female *n* (%) Gp1 59 (59.6); Gp2 65 (63.1) Mean age, years ± SD Gp1 74.1 ± 7.6; Gp2 71.4 ± 9.3 Range Gp1 53–93; Gp2 48–95 Baseline MMSE, mean ± SD Gp1 18.8 ± 4.8; Gp2 18.5 ± 4.8	Primary outcomes: MMSE; ADAS-cog; DAD and NPI Methods of assessing outcomes: Patients who exhibited decline or no change from baseline on the MMSE and whose physician was not sufficiently certain of clinical benefit to warrant continued treatment were rated as showing 'no apparent clinical benefit' and were randomised into the double-blind phase. Patients were assessed at baseline (week 0 of double-blind phase), and at weeks 6 and 12. Length of follow-up: 12 weeks Also a 12 week single-blind phase after RCT – not data extracted

Results

Outcomes	Donepezil 10 mg/day (*n* = 99)	Placebo (*n* = 103)	*p*-Value
MMSE [least-square mean change] at week 12 (OC) (estimated from figure)	1.7	0.5	$p < 0.05$ versus placebo washout

At week 12, there was a significant difference in favour of patients who received continuous donepezil treatment, over those who switched to placebo, on the MMSE (treatment difference –1.13; $p < 0.05$)

Treatment differences at week 12	Donepezil 10 mg/day versus placebo	
ADAS-cog	0.57	$p = 0.53$
DAD	–3.67	$p = 0.11$
NPI	3.16	$p < 0.05$

Comments:
Outcomes on the ADAS-cog showed differences favouring treatment with continuous donepezil over placebo. Differences in favour of continuous donepezil treatment over placebo were seen on the DAD.
Significant benefits in patients receiving continuous donepezil treatment versus placebo were also seen on NPI.
Patients receiving continuous donepezil treatment showed benefits in significantly more domains, compared with patients who underwent placebo washout

continued

Adverse effects

Comments: Reports that during the double-blind phase, the number of adverse events was low. No further information provided.

Methodological comments

- Allocation to treatment groups: Patients rated as showing 'no apparent clinical benefit' at weeks 12, 18 or 24 were randomised into the double-blind phase of the study. No further details provided.
- Blinding: Randomised part of study described as double-blind, placebo controlled. No further information provided.
- Comparability of treatment groups: There was no difference in patient characteristics between the treatment groups. However, the study does not state whether baseline characteristics were taken at the beginning of the open-label phase or at the randomisation stage.
- Method of data analysis: Results were for the intention-to-treat (ITT) population (patients receiving at least 1 dose of medication and who provided at least 1 post-baseline assessment), week 12 observed cases. Efficacy variables were analysed by an ANCOVA model, with terms for country and baseline. Results were reported as least-squares mean change from baseline. Last observation carried forward (LOCF) and observed cases analyses were implemented.
- Sample size/power calculation: Not reported
- Attrition/dropout: None reported

General comments

- Generalisability: Patients with mild to moderate possible or probable AD. Patients likely to be skewed population as many dropouts before randomisation; criteria to judge 'no benefit' were largely subjective; contamination as 15 patients were responders.
- Outcome measures: Outcome measures relevant to study area and measured appropriately
- Inter-centre variability: None reported
- Conflict of interests: Three authors were employed by Pfizer/Eisai Inc.

Quality criteria for Nunez *et al.*[48]

1. Was the assignment to the treatment groups really random?	Unknown
2. Was the treatment allocation concealed?	Unknown
3. Were the groups similar at baseline in terms of prognostic factors?	Reported
4. Were the eligibility criteria specified?	Adequate
5. Were outcome assessors blinded to the treatment allocation?	Unknown
6. Was the care provider blinded?	Partial
7. Was the patient blinded?	Partial
8. Were the point estimates and measure of variability presented for the primary outcome measure?	Inadequate
9. Did the analyses include an intention-to-treat analysis?	Inadequate
10. Were withdrawals and dropouts completely described?	Unknown

Reference and design	Intervention	Participants	Outcome measures
Author: Rogers *et al*.[53] Year: 1996 Country: USA Study design: Multicentre, double-blind, parallel group, placebo controlled, randomised trial Number of centres: Not reported (in any of the papers). There are 10 listed members of the donepezil study group, all at different locations Funding: Eisai America Inc (Teaneck, NJ, USA) and Eisai Co Ltd (Tokyo, Japan)	Treatment arms: 4 Groups (i) 1 mg donepezil (ii) 3 mg donepezil (iii) 5 mg donepezil (vi) placebo Medication (film-coated tablet) taken once daily for 12 weeks in the evening. The double-blind phase of the study was followed by a 2 week single-blind washout phase. Patients were supervised for 4 hours at the study site after receiving the 1st dose of medication Other interventions used: The protocol indicates that some patients may have already been receiving treatment with sympathomimetic amines or antihistamines and that if so, this medication was interrupted for at least 48 hours prior to clinic visits	Number of participants: 161 patients were randomised to treatment Gp1 1 mg donepezil $n = 42$ Gp2 3 mg donepezil $n = 40$ Gp3 5 mg donepezil $n = 39$ Gp4 placebo $n = 40$ Sample attrition/dropout: 20 patients withdrew/failed to complete Gp1 1 mg donepezil $n = 8$ Gp2 3 mg donepezil $n = 2$ Gp3 5 mg donepezil $n = 5$ Gp4 placebo $n = 5$ (see results for reasons) Sample crossovers: There was no crossover in this study. Inclusion/exclusion criteria for study entry: Age between 55 and 85 years Diagnosis of mild to moderately severe AD (DSM-III-R and NINCDS criteria) made at least 1 year prior to entry. AD diagnosis supported by computerised tomographic or MRI studies performed during 6 months prior to study entry MMSE 10–26 inclusive and CDR rating of 1 or 2. Fully ambulatory or able to walk with cane or walker. Vision and hearing sufficient for compliance with testing. Female patients 2 years postmenopausal or surgically sterilised prior to study entry Exclusion criteria Evidence of other psychiatric or neurological disorder. Clinically significant or active gastrointestinal, renal, hepatic, endocrine or cardiovascular diseases. Diabetes (any form), obstructive pulmonary disease, haematologic or oncologic disorders of recent onset (within 2 years), vitamin B_{12} or folate deficiency. History of alcohol or drug abuse. Known hypersensitivity to cholinesterase inhibitors. Receipt of other investigational drugs within one month of trial entry Characteristics of participants: Gender: Male/Female Gp1: 13/29; Gp2: 18/22; Gp3: 14/25; Gp4: 19/21 Race: Caucasian/Black/Other Gp1: 41/1/0; Gp2: 37/3/0; Gp3: 37/1/1; Gp4: 40/0/0 Age, years: Mean (range) Gp1: 72.6 (55–85); Gp2: 71.0 (54–85); Gp3: 72.9 (55–85); Gp4: 70.6 (56–84) Weight, kg: Mean (range) Gp1: 61.9 (32.0–86.4) $p = 0.03$; Gp2: 68.5 (44.0–99.5); Gp3: 64.4 (42.7–88.1); Gp4: 70.2 (45.5–104.3) Height, cm: Mean (range) Gp1: 162.4 (132.0–185.0) $p = 0.03$; Gp2: 168.2 (149.9–210.0); Gp3: 161.7 130.0–177.8); Gp4: 165.1 (129.5–193.0)	Primary outcomes: Change from baseline score for ADAS-cog. Endpoint rating of the investigator's CGIC Secondary outcomes: Change from baseline scores in: Uniform activities of daily living (ADL) MMSE CDR-SB (the sum of ratings for 6 CDR functional domains) QoL-P, QoL-C (to assess well-being) Methods of assessing outcomes: CGIC ratings were based on the investigator's total experience with the subject and were performed by clinicians who were unaware of the patients' performance on psychometric rating scales. ADL conducted by interviewing the caregiver MMSE conducted with the subject QoL-P assessed by patient QoL-C assess by caregiver Therapeutic drug monitoring also took place, measuring plasma concentrations of donepezil. AchE activity was measured in RBC membranes (data not extracted as per protocol) Safety All adverse events reported by patients or noticed by caregivers or physicians were recorded, together with date of onset and cessation, severity and relationship to trial medication Haematology, urinalysis and vital signs were also performed at each visit (not extracted)

continued

Reference and design	Intervention	Participants	Outcome measures
		Small but statistically significant differences were found between the four groups of patients with respect to body weight and height. These differences were not sufficient to influence the outcome of the trial. Baseline values for outcome measures [mean (min, max)] ADAS-cog: Gp 1 26.6 (12.2, 47.0) Gp2 29.2 (7.7, 54.7) Gp3 29.1 (12.0, 59.0) Gp4 27.2 (9.7, 49.3) ADL: Gp 1 94.7 (63, 248) Gp2 98.8 (62, 210) Gp3 105.5 (62, 255) Gp4 92.4 (63, 271) MMSE: Gp 1 19.6 (10.0, 28.0) Gp2 18.6 (9.0, 26.0) Gp3 18.0 (6.0, 26.0) Gp4 18.2 (8, 27) QoL-P: Gp 1 291.1 (221, 350) Gp2 270.3 (155, 350) Gp3 283.7 (165, 350) Gp4 295.4 (187, 350) QoL-C: Gp 1 251.2 (83, 346) Gp2 251.0 (160, 340) Gp3 251.1 (125, 350) Gp4 249.1 (80, 340) CDR-SB: Gp 1 6.6 (3.5, 12.0) Gp2 6.9 (3.5, 14.0) Gp3 7.3 (3.5, 13.0) Gp4 6.7 (4.0, 13.0)	Length of follow-up: This was a 14 week study – 12 weeks treatment and 2 weeks washout. Separate publications describe the open label extension which some patients participated in for up to 240 weeks Patients were asked to fast for at least 2 hours prior to clinic visits Preexisting treatment with sympathomimetic amines or antihistamines was interrupted for at least 48 hours prior to clinic visits. Assessments occurred at screening visit, at baseline, at 1, 3, 6, 9, 12 and 14 weeks after the start of treatment

Results

Outcomes		Donepezil 1 mg $n = 42$	Donepezil 3 mg $n = 40$	Donepezil 5 mg $n = 39$	Placebo $n = 40$	p-Value
ADAS-cog: Mean change from baseline ± SE Values in *italics* estimated from graph	Week 1	*–0.8 ± 0.6*	*–1.0*	*–1.9*	*–0.5 ± 0.6*	All p-values are relative to placebo unless otherwise stated
	Week 3	*–2.0 ± 0.8*	*–3.1* ($p < 0.04$)	*–3.8* ($p < 0.01$)	*–0.9 ± 0.8*	
	Week 6	*–2.0 ± 0.7*	*–2.4*	*–2.2*	*–0.5 ± 0.8*	
	Week 9	*–1.5 ± 0.8*	*–1.3*	*–3.0* ($p < 0.007$)	*0.0 ± 0.9*	
	Week 12	*–1.2 ± 0.7*	*–1.4* ($p < 0.036$)	*–2.5* ($p < 0.002$)	*0.8 ± 0.8*	
	Endpoint (ITT)	*–0.9 ± 0.7*	*–1.4*	*–2.5* ($p < 0.003$)	*0.7 ± 0.7*	
	Washout period	*1.1 ± 0.7*	*0.3*	*–2.1* ($p < 0.001$)	*1.6 ± 0.8*	

Comments: SE for 3 mg and 5 mg data not visible on graph, unclear whether this is because SEs are too small to be seen (unlikely) or whether have been deliberately left off the graph

					p-Value for dose response analysis (dose trend)
ADAS-cog: ADJUSTED Mean change from baseline at endpoint (ITT) mean (min, max)	–0.9 (–11.3, 12.0) not sig diff to placebo $p = 0.105$	–1.4 (–12.0, 11.0) relative to placebo $p = 0.036$	–2.5 (–8.0, 7.0) relative to placebo $p = 0.002$	0.7 (–7.0, 14.5)	0.0359 difference favours donepezil

ADAS-cog comments: The effect on ADAS-cog was related to time and dose, the dose; trend was statistically significant. The ADAS-cog response was maintained at the end of the 2-week washout in the 5-mg group (–2.0), but not in the 3-mg group (0.3).

continued

Comments: The paper also gives the baseline mean (min, max) values for the above ADAS-cog scores. The min, max values presented here are obviously not the min change and max change but rather the natural min, and max values for each score at endpoint

CGIC at week 12 and endpoint	**Donepezil 1 mg *n* = 42**	**Donepezil 3 mg *n* = 40**	**Donepezil 5 mg *n* = 39**	**Placebo *n* = 40**
Week12 success/failure	29 (83%)/6 (17%)	32 (87%)/5 (14%)	32 (89%)/4 (11%)	29 (81%)/7 (20%)
Endpoint success/failure	34 (82%)/7 (18%)	33 (83%)/7 (18%)	34 (90%)/4 (11%)	32 (80%)/8 (20%)

CGIC comments: The authors state that the short duration of the study complicated interpretation of the CGIC results. A 12-week period is one considered to represent a stable window of time relative to the progression of AD. Thus, in the majority of patients the condition was scored as being unchanged. Assessing this outcome in terms of treatment failures (CGIC scores 5–7) or treatment successes (CGIC scores 1–4) however demonstrated that treatment failure decreased in a dose dependent manner (i.e. failure rate of 11% for 5-mg dose compared to 20% for placebo, was significantly lower – $p = 0.039$ with the 5 mg dose). Note: endpoint values should be ITT analysis, but for 1 mg total only comes to 41 and for 5 mg total is only 38, so 1 patient missing from each of these groups

	Donepezil 1 mg *n* = 42	**Donepezil 3 mg *n* = 40**	**Donepezil 5 mg *n* = 39**	**Placebo *n* = 40**	***p*-Value for dose response analysis**
ADL: Adjusted mean change from baseline at Endpoint (ITT) mean (min, max)	4.0 (–25, 97)	0.6 (–21, 30)	–3.1 (–36, 15)	1.5 (–38, 57)	0.0684 difference favours donepezil (NS)
ADL comments: At endpoint adjusted mean ADL scores showed no effect for the placebo, 1- and 3-mg dose groups relative to baseline. The 5-mg group mean score was improved relative to baseline and the difference between the 1-mg (4.0) and 5-mg (–3.1) dose groups demonstrated a trend to significance ($p = 0.068$)					
MMSE: Adjusted Mean change from baseline at Endpoint (ITT) mean (min, max)	0.6 (–4.0, 7.0)	0.9 (–7.0, 5.0)	2.0 (–1.0, 7.0)*** *** = $p < 0.05$ relative to 1 mg	1.2 (–6.0, 8.0)	0.0275 difference favours donepezil
MMSE: Mean change from baseline scores Estimated from figure					Note: fig. did say ± SE but no error bars were visible
Week 1	*0.5*	*0.5*	*1.05*	*1.05*	
Week 3	*0.65*	*0.65*	*1.40*	*1.00*	
Week 6	*1.15*	*1.30*	*1.80*	*0.81*	
Week 9	*0.11*	*0.34*	*2.29*	*0.39*	
Week 12	*0.50*	*0.80*	*2.00*	*1.00*	
Endpoint (ITT)	*0.66*	*0.80*	*2.00*	*1.37*	
Washout period	*0.30*	*0.70*	*1.55*	*0.4*	
MMSE comments: Mean MMSE scores for the placebo, 1- and 3-mg dose groups were similar but the 5-mg group had greater improvement in mean change from baseline at every post baseline visit and a statistically significant dose–response relationship was found. At endpoint, the *adjusted* mean change in MMSE in the 5-mg group (2.0) was significantly greater than that seen in the 1-mg group (0.6; $p < 0.05$)					
QoL-P: Adjusted mean change from baseline at Endpoint (ITT) mean (min, max)	0.7 (–90, 60)	2.6 (–90, 100)	8.8 (–143, 110)	–1.3 (–74, 78)	0.0369 difference favours donepezil
QoL-C: Adjusted mean change from baseline at Endpoint (ITT) mean (min, max)	–5.3 (–120, 74)	0.0 (–70, 97)	0.3 (–120, 124)	3.7 (–50, 140)	0.8860 difference favours placebo (NS)
QoL-P and QoL-C Comments: QoL-P scores showed extensive inter- and intra-patient variability and the pairwise differences between the treatment groups were not statistically significant. The dose trend analysis showed statistically significant improvement ($p < 0.05$). The QoL-C also showed marked inter-subject variability with no statistical evidence of improvement over placebo in any of the treatment group, suggesting that caregivers may not be useful informants about the patient's inner feelings. Well-being was assessed for relationships, eating and sleeping, and social/leisure activity					

continued

CDR-SB: Adjusted mean change from baseline at Endpoint (ITT) mean (min, max)	0.18 (–2.0, 5.0)	0.23 (–3.0, 6.0)	–0.11 (–2.0, 3.0)	0.10 (–2.0, 3.0)	0.3375 difference favours donepezil (NS)
Mean change from baseline in CDR-SB					
Week 1	–0.07	–0.07	–0.07	–0.13	
Week 3	–0.09	–0.16	–0.07	–0.09	
Week 6	–0.05	0.00	–0.04	0.09	
Week 9	–0.12	0.02	–0.17	0.09	
Week 12	0.18	0.23	–0.11	0.10	
Endpoint (ITT)	0.10	0.04	–0.15	0.04	
Washout period	0.23	0.40	–0.03	0.19	

CDR-SB comments: No statistically significant effects of donepezil and adjusted mean change from baseline of the CDR-SB scores were found, and mean scores for the placebo, 1-, and 3-mg groups were similar. However the 5-mg group showed greater improvement than placebo for all visits after week 3.

General comments: The paper also gives the baseline mean (min, max) values for the above measures. The min, max values presented here are obviously not the min change and max change but rather the natural min and max values for each score at endpoint. Although not implicitly stated it seems likely that only the endpoint values are ITT; the other weekly measures are not ITT and the numbers contributing to these outcomes are not known.

Withdrawals total number	8	2	5	5
Reasons				
Adverse event(s)	5	2	3	2
Protocol violation	1	0	1	3
Withdrew consent	1	0	0	0
Request of patient/investigator	1	0	1	0
Other	0	0	0	0

Comments: Paper does not indicate how many patients had withdrawn from each group by each clinic visit.

Adverse events					Assume not tested statistically
Total no of patients with 1 or more adverse events (%)	27 (64)	27 (68)	26 (67)	26 (65)	
Gastrointestinal					
Nausea/vomiting	3 (7)	0	4 (10)	2 (5)	
Diarrhoea	0	1 (3)	4 (10)	1 (3)	
Gastric upset	0	2 (5)	3 (8)	2 (5)	
Constipation	1 (2)	2 (5)	3 (8)	1 (3)	
Other events					
Dizziness	2 (5)	1 (3)	3 (8)	4 (10)	
Nasal congestion	1 (2)	5 (13)	2 (5)	3 (8)	
Common cold	4 (10)	2 (5)	2 (5)	2 (5)	
Headache	4 (10)	2 (5)	1 (3)	3 (8)	
Flushing	4 (10)	1 (3)	1 (3)	1 (3)	
Agitation	3 (7)	2 (5)	1 (3)	2 (5)	
Urinary tract infection	1 (2)	3 (8)	1 (3)	2 (5)	
Coughing	1 (2)	4 (10)	1 (3)	2 (5)	
Accident	1 (2)	1 (3)	4 (10)	1 (3)	
Pain	3 (7)	1 (3)	2 (5)	1 (3)	

Comments: Adverse events were defined as events that began during or after administration of the first dose of study medication or became more severe during treatment. The most frequently encountered adverse events at the 5-mg dosage compared with placebo were nausea/vomiting, diarrhoea, dizziness, gastric upset and constipation. The majority of treatment emergent adverse events were of mild to moderate intensity and in most cases there was no apparent relationship to the dose of donepezil. Donepezil had no clinically significant effect on vital signs, haematology or clinical biochemistry test. Importantly, donepezil was not associated with any hepatotoxicity in this study.

continued

Methodological comments

- Allocation to treatment groups: States 'allocation by randomisation' but no further details given.
- Blinding: All tablets film coated: presumably this is mentioned to indicate that all the tablets were indistinguishable from one another although this is not stated outright. Treatment phase is described as double-blind but no further details regarding this are given. The placebo washout phase was only single-blind.
- Comparability of treatment groups: The paper states that small but statistically significant differences were found between the four groups of patients with respect to body weight and height and goes on to add that these differences were not sufficient to influence the outcome of the trial. All groups were predominantly Caucasian.
- Method of data analysis: Efficacy and safety analyses were performed on an ITT population which included all patients who were randomised to treatment, received at least one dose of study drug, and had at least one postbaseline data point. The efficacy conclusions were based on the combined results at each patient's last assessment during double-blind therapy – defined as the study endpoint. All hypothesis tests were two-sided, and p-values ≤ 0.05 were considered to be statistically significant. On more than one occasion graph legends indicated mean±SE was displayed but this was not visible either for some or all of the groups. This may be because the SE was too small to be displayed on the graph but this seems unlikely; no SEs are reported in the text in these instances. For some results the mean change is given with the range of endpoint values rather than the spread about the mean change. ANCOVA used to examine differences in efficacy across the four treatment groups. Where differences were observed, pairwise comparisons using Fischer's 2-tailed least significant difference procedure were undertaken, except for CGIC scores where Cochran–Mantel–Haenszel test was used. Bonferroni adjustments used to accommodate comparisons of multiple dose groups.
- Sample size/power calculation: The planned study population of 40 patients per group was based on review of clinical studies of other ChE inhibitors and previous Phase I studies conducted using donepezil. The sample size was intended to achieve 80% power to detect a 2.5 point improvement in ADAS-cog with $p \leq 0.05$ for donepezil treatment compared to placebo.
- Attrition/dropout: Withdrawals were described although the timepoint at which these occurred was not given so there is no way of working out how many patients' data contributed to the outcomes at the intermediate timepoints of the study.

General comments

- Generalisability: The criteria limited potential participants to those without significant concomitant disease so the results may not be generalisable to a large number of AD patients who are quite likely to have other diseases.
- Outcome measures: Patients were asked to fast for at least 2 hours prior to clinic visits. Preexisting treatment with sympathomimetic amines or antihistamines was interrupted for at least 48 hours prior to clinic visits. Measures seemed relevant and measured appropriately. QoL measures probably not valid. No description of what the scores mean was given.
- Inter-centre variability: Study procedures were undertaken in the same sequence for each centre, e.g. psychometric and neurological examination, physical examination, vital signs, collection of blood samples. Paper states CGIC scores were analysed using Cochran–Mantel–Haenszel test which 'included adjustments for centre differences', indicating that there may well have been some variability but this is not discussed further.
- Conflict of interests: Eisai America Inc. (Teaneck, NJ, USA) and Eisai Co Ltd (Tokyo, Japan).

Quality criteria for Rogers *et al.*[53]

1. Was the assignment to the treatment groups really random?	Unknown
2. Was the treatment allocation concealed?	Unknown
3. Were the groups similar at baseline in terms of prognostic factors?	Reported
4. Were the eligibility criteria specified?	Adequate
5. Were outcome assessors blinded to the treatment allocation?	Unknown
6. Was the care provider blinded?	Partial
7. Was the patient blinded?	Partial
8. Were the point estimates and measure of variability presented for the primary outcome measure?	Partial
9. Did the analyses include an intention-to-treat analysis?	Inadequate
10. Were withdrawals and dropouts completely described?	Adequate

Reference and design	Intervention	Participants	Outcome measures
Author: Rogers *et al*.[51] Year: 1998 Country: USA Study design: Multicentre RCT Number of centres: 20 Funding: Eisai Inc. (USA) and Eisai Co. Ltd. (Japan)	Treatment arms: (including dosage, length of treatment) (1) 5 mg/day donepezil (don5) (2) 10 mg/day donepezil (don10) (3) Placebo (pl) 24 weeks' treatment followed by 6-week single-blind placebo washout. A single dose was given to all groups each evening. A blinded forced titration scheme was used for don10 pts. They received 5 mg/day for the first week and 10 mg/day for the remainder of the study. Other interventions used: Concomitant medications such as anticholinergics, anticonvulsants, antidepressants and antipsychotics were not allowed during the study. Drugs with CNS activity were either prohibited or partially restricted. All other medications were permitted	Number of participants: Don5 $n = 154$ Don10 $n = 157$ Pl $n = 162$ Sample attrition/dropout: Completion rates: 85% (don5), 68% (don10), 80% (placebo) Discontinuation due to adverse events: 6% (don5), 16% (don10), 7% (pl) Sample crossovers: none Inclusion/exclusion criteria for study entry: Inclusion: diagnosed with 'uncomplicated AD'; ≥ 50 years old; probable AD by NINCDS-ADRDA criteria, with pts also fitting DSM-III-R categories of 290.00 or 290.10, with no clinical or laboratory evidence of a cause other than AD for their dementia; MMSE 10–26; Clinical Dementia Rating of 1 (mild) or 2 (moderate) at both screening and baseline. Patients were required to have a reliable caregiver Exclusions: pts with insulin-dependent diabetes mellitus or other endocrine disorders; asthma or obstructive pulmonary disease; clinically significant uncontrolled gastrointestinal, hepatic or cardiovascular diseases. Patients who were known to be hypersensitive to ChE inhibitors or had been taking tacrine and/or other investigational medications within one month of baseline Characteristics of participants: Characteristic / don5 / don10 / pl Age* (year): 72.9 ± 0.6 / 74.6 ± 0.6† / 72.6 ± 0.6 Age range: 51–86 / 53–94 / 56–88 Sex M (%): 57 (37) / 60 (38) / 63 (39) F (%): 97 (63) / 97 (62) / 99 (61) Race white (%): 146 (95) / 150 (96) / 153 (94) African-American (%): 5 (3) / 3 (2) / 6 (4) Other (%): 3 (2) / 4 (3) / 2 (3) Screening CDR 0.5 (%): 1 (1)‡ / 0 / 0 1.0 (%): 114 (74) / 119 (76) / 121 (75) 2.0 (%): 39 (25) / 37 (24) / 41 (25) Screening MMSE*: 19.0 ± 0.4 / 18.9 ± 0.4 / 19.2 ± 0.4 * Values are means ± SEM † Difference in mean age between donepezil 10 and placebo was significant, $p = 0.03$ ‡ Patient was subsequently excluded as a protocol violation	Primary outcomes: ADAS-cog; CIBIC-plus. For CIBIC-plus, the order of interviewees (patient and caregiver) was randomised at each visit Secondary outcomes: MMSE; Clinical Dementia Rating-Sum of the Boxes (CDR-SB) – ratings agreed by patient's assessment team (excluding CIBIC-plus clinician); patient rated QoL. Donepezil concentrations in blood plasma and an analysis quantifying inhibition of red blood cell AChE activity in blood samples (not data extracted as not per protocol). Outcomes assessed at baseline and at 6-week intervals. Outcomes also reported at 30 weeks, to include the washout phase (not data extracted as not per protocol) Methods of assessing outcomes: not stated Length of follow-up: 24 weeks plus 6-week placebo washout

Abbreviations: pts, patients; pl, placebo group; don5, 5 mg/day donepezil group; don10, 10 mg/day donepezil group

continued

Results

ADAS-cog mean change from baseline	**Don5**	**Don10**	**Placebo**	***p*-Value Don5/don10**
6 weeks (estimated from figure)	–1.9	–1.6	–1.3	
12 weeks (estimated from figure)	–1.3	–2.0	0.75	0.0007/<0.0001
18 weeks (estimated from figure)	–0.8	–1.4	1.4	0.00012/<0.0001
Endpoint (24 weeks) (figures stated)	–0.67	–1.06	1.82	<0.0001/<0.0001

Comments
ADAS-cog: 11-item scale, with score range 0–70. Lower scores indicate lesser severity, negative change indicates clinical improvement.
CIBIC-plus: 7 point Likert-type scale used for scoring, where 1 = marked improvement, 4 = no change, 7 = marked worsening.
ADAS-cog: figs estimated from Figure 1, endpoint figs from text.

ADAS-cog (ITT-LOCF)	**Don5**	**Don10**	**Placebo**	***p*-Value[‡] Don5/don10**
Patient numbers	*n* = 152	*n* = 150	*n* = 153	
Mean baseline score*[†]	26.28 ± 0.96	27.41 ± 0.86	27.28 ± 0.96	
Endpoint: mean ADAS-cog change from baseline*	–0.67 ± 0.51	–1.06 ± 0.51	1.82 ± 0.49	
Drug – placebo difference	–2.49	–2.88		<0.0001/<0.0001
Mean change at 30 weeks*[§]	2.29 ± 0.56	2.96 ± 0.64	2.91 ± 0.57	

CIBIC-plus (ITT-LOCF)	**Don5**	**Don10**	**Placebo**	***p*-Value[‡] Don5/don10**
Patient numbers	*n* = 149	*n* = 149	*n* = 152	
Mean baseline score*[†]	–	–	–	
CIBIC-plus value at endpoint*	4.15 ± 0.09	4.07 ± 0.07	4.51 ± 0.08	
Drug – placebo difference	0.36	0.44		0.0047/<0.0001

MMSE (ITT-LOCF)	**Don5**	**Don10**	**Placebo**	***p*-Value Don5/don10**
Patient numbers	*n* = 153	*n* = 150	*n* = 154	
Mean baseline score*[†]	19.44 ± 0.38	19.17 ± 0.37	19.40 ± 0.37	
Endpoint: mean change from baseline*	0.24 ± 0.29	0.39 ± 0.29	–0.97 ± 0.28	
Drug – placebo difference	1.21	1.36		0.0007/0.0002

CDR-SB (ITT-LOCF)	**Don5**	**Don10**	**Placebo**	***p*-Value Don5/don10**
Patient numbers	*n* = 154	*n* = 151	*n* = 153	
Mean baseline score*[†]	7.11 ± 0.19	7.13 ± 0.19	6.98 ± 0.19	
Endpoint: mean change from baseline*	–0.01 ± 0.14	–0.02 ± 0.14	0.58 ± 0.14	
Drug – placebo difference	0.59	0.60		0.0008/0.0007

Comments:
* Values are means ± SEM.
[†] Mean baseline scores at randomisation.
[‡] Despite the difference in age between the groups, the treatment by age interaction was not found to be statistically significant. An ANCOVA model where response = overall means + baseline score + age at baseline + treatment effect + site effect + random effect was used as the primary model to test for overall treatment effect using type III sums of squares.
[§] Means are the change from baseline at 30 weeks after a 6-week single-blind washout.

continued

Change in ADAS-cog scores from baseline	Don5	Don10	Placebo
≥ 7 points	15.4%	25.2%	7.8%
≥ 4 points	37.8%	53.5%	26.8%
≥ 0 points	79.7%	81.1%	57.7%

Comments:
The % of patients with poorer ADAS-cog scores in the endpoint analysis relative to baseline were: 42.3% (pl), 20.3% (don5), 18.9% (don10). Therefore at least 80% of patients receiving donepezil did not experience cognitive worsening, compared with 57.7% of placebo patients over the 24 weeks of treatment.

Mean CIBIC plus score (estimated from figure)	Don5	Don10	Placebo	*p*-Value Don5/don10
Week 6	3.95	3.95	3.95	
Week 12	3.98	3.9	4.2	0.0157/0.009
Week 18	4.1	3.98	4.4	0.0244/0.0002
Endpoint	4.15	4.05	4.5	0.0047/<0.0001

Comments:
The differences in mean drug-placebo CIBIC-plus scores at endpoint were dose dependent at 0.36 for don5 and 0.44 for don10 groups. Only 11% of placebo patients, as compared with 26% of don5 group and 25% of don10 patients, were scored as improved (CIBIC-plus ≤ 3). Overall, donepezil increased the number of treatment successes (CIBIC-plus ≤ 4) and reduced the number of treatment failures (CIBIC-plus ≥ 5; $p = 0.0018$). The % of patients who failed visits at least half the time was 45% in placebo patients, 33% in don5 patients and 25% in don10 patients.

Mean MMSE change from baseline (estimated from figure	Don5	Don10	Placebo	*p*-Value Don5/don10
6 weeks	0.3	0.4	0	
12 weeks	0.75	1	–0.5	0.0002/<0.0001
18 weeks	0.5	0.9	–0.75	0.0006/<0.0001
Endpoint	0.2	0.4	–1.0	0.0007/0.0002

Comments:
Numbers estimated from Fig 4.

Mean CDR-SB change from baseline (estimated from figure)	Don5	Don10	Placebo	*p*-Value Don5/don10
6 weeks	–0.1	–0.05	0.01	
12 weeks	–0.2	–0.04	0.1	
18 weeks	–0.15	–0.06	0.3	0.0105/0.0337
Endpoint	0	0	0.55	0.0007/0.0008

Comments

Mean patient-rated QoL score change from baseline (estimated from figure)	Don5	Don10	Placebo	*p*-Value
6 weeks	5	–1	0	
12 weeks	8	2.5	–0.25	
18 weeks	7	3.0	–0.2	
Endpoint	11*	7.5	–2	*$p = 0.05$ at week 24

Comments:
In text: The 5 mg/day dose group achieved significant improvement at week 24 ($p = 0.05$). "Significant differences were not evident at the study endpoint." Unclear, seems to contradict as week 24 is the endpoint?

continued

Adverse effects – number (%) patients with symptoms	Don5 (n = 154)	Don10 (n = 157)	Placebo (n = 162)
Fatigue	8 (5)	12 (8)†	3 (2)
Diarrhoea	14 (9)	27 (17)†	11 (7)
Nausea	6 (4)	26 (17)†	6 (4)
Vomiting	5 (3)	16 (10)†	3 (2)
Anorexia	3 (2)	11 (7)	3 (2)
Muscle cramps	9 (6)	12 (8)†	1 (1)
Dizziness	15 (10)	13 (8)†	7 (4)
Rhinitis	1 (1)	9 (6)	4 (2)

Comments:
† $p \leq 0.05$. Overall p-values were calculated only for preferred terms where the overall incidence rate was ≥ 5%.
Most of these events were of mild severity, although nausea and vomiting were occasionally of moderate severity. The higher incidence of cholinergic side effects experienced in the don10 group was due to the forced, rapid titration schedule used. One placebo patient and one don10 patient died during the study; their deaths were not related to the study drug. 31 patients (6%) experienced one or more adverse events during the study or within 1 month of its termination, with most considered unrelated to study drug. Slightly more patients experienced serious adverse effects in the don10 group (15 patients; 10%) than in the don5 (7 patients; 5%) or placebo (9 patients; 6%) groups. The % of adverse events judged as possibly related to treatment was lowest for the don10 group. No events were judged probably or definitely related to treatment.

Dose	Serious adverse event	Relationship to drug
Placebo	Ischaemia myocardial,‡ syncope	Possibly related, but sponsor judged event to be not related
	Embolus pulmonary†	Not related, but sponsor judged event to be possibly related
	Abdominal disturbance,† gastrointestinal disorder†	Possibly related
	Bronchitis	Possibly related, but sponsor judged event to be not related
5 mg/day	Angina pectoris	Possibly related, but sponsor judged event to be not related
	Premature ventricular contractions,† syncope,† dizziness†	Possibly related, but sponsor judged event to be not related
	Infection pyelonephritis,† renal failure†	Possibly related, but sponsor judged event to be not related
10 mg/day	Head pressure,† blood pressure oscillatory,† drooling,† ataxia,† dysarthria†	Possibly related
	Agitation†	Possibly related, but sponsor judged event to be not related
	Accident, fracture bone, hypoxia	Possibly related
	Nausea, vomiting, dehydration, thrombosis venous deep	Possibly related

Comments:
Table 4 lists serious adverse events. Only those considered to be possibly related to treatment are included here.
† Patient withdrew because of this serious adverse event
‡ Patient withdrew because of myocardial ischaemia and two non-serious adverse events: movement disorder and psychosis.

Methodological comments

- Allocation to treatment groups: Assigned by a computerised randomisation schedule.
- Blinding: The 10 mg/day dosage group received a blinded forced titration scheme. Investigator assessment of relationship to treatment for all adverse events, serious and non-serious, was conducted under blinded conditions. The trial was described as double-blind, but no further details are given.
- Comparability of treatment groups: Patient demographic characteristics did not differ between treatment groups, except for age. The mean age of the donepezil 10 mg/day group was 2 years older than the mean for the placebo group (p = 0.03). Other characteristics were similar between groups (data not shown). However, despite this, the treatment by age interaction in the ITT-LOCF analyses was not found to be statistically significant.
- Method of data analysis: (1) 'fully evaluable population' analysis for all patients who completed 24 weeks of double-blind treatment with at least 80% compliance at week 24 and had at least 2 other visits during double-blind phase with no significant protocol violations. (2) ITT, including all patients who were randomised to treatment, received at least 1 dose, provided complete baseline data and a minimum of 1 post-baseline data point. Efficacy conclusions were based on results

continued

at each patient's last assessment, defined as study endpoint (i.e. LOCF). For continuous efficacy variables, changes from baseline were measured using ANCOVA. In cases where differences existed, pairwise comparisons between active treatment and placebo were undertaken using Fisher's 2 tailed least significant difference procedure. The categorical efficacy variable (CIBIC-plus) was analysed using the Cochran–Mantel–Haenszel (C-M-H) test. ANOVA and C-M-H methods were used to assess comparability of the three groups for continuous and categorical demographic variables, respectively. Intragroup changes in vital signs were analysed using paired *t*-tests, and ANOVA was used to detect differences between treatments. Point estimates and interquartile ranges were presented in box and whisker plots of changes from baseline through time. Means ± SEM were presented for outcome measures. All hypothesis tests were 2 sided, with analyses being significant if a ≤ 0.05 level was achieved.

- Sample size/power calculation: Not stated
- Attrition/dropout: 20% (placebo), 15% (5 mg/day donepezil), 32% (10 mg/day donepezil).

General comments

- Generalisability: The study was limited to patients diagnosed with uncomplicated probable AD; ≥50 years old; fitting DSM-III-R categories of 290.00 or 290.10, with no clinical or laboratory evidence of a cause other than AD for their dementia; an MMSE score of 10–26; a Clinical Dementia Rating of 1 (mild) or 2 (moderate) at both screening and baseline. Patients with certain concomitant diseases were excluded (see above).
- Outcome measures: Relevant outcome measures were used and these were measured appropriately
- Inter-centre variability: ANCOVA model included a factor for treatment-by-centre interaction. C-M-H test included an adjustment for centre differences.
- Conflict of interests: Two authors work for Eisai Inc, USA, who funded this study (together with Eisai Co. Ltd, Japan) and manufacture donepezil.

Quality criteria for Rogers *et al.*[51]

1. Was the assignment to the treatment groups really random?	Adequate
2. Was the treatment allocation concealed?	Unknown
3. Were the groups similar at baseline in terms of prognostic factors?	Reported (age significantly higher for high dose group than placebo)
4. Were the eligibility criteria specified?	Adequate
5. Were outcome assessors blinded to the treatment allocation?	Partial
6. Was the care provider blinded?	Partial
7. Was the patient blinded?	Partial
8. Were the point estimates and measure of variability presented for the primary outcome measure?	Adequate
9. Did the analyses include an intention-to-treat analysis?	Inadequate
10. Were withdrawals and dropouts completely described?	Partial

Reference and design	Intervention	Participants	Outcome measures
Author: Rogers *et al*.[52] Year: 1998 Country: USA Study design: RCT, multicentre Number of centres: 23 Funding: Eisai Inc and Eisai Co Ltd	Treatment arms: 12 weeks of treatment with: (1) 5 mg donepezil (2) 10 mg donepezil (all had a blinded forced titration phase to minimise the likelihood of reactions to acute acetylcholine inhibition where 5 mg was given for the first 7 days) (3) placebo All administered once daily at bedtime. Each dose consisted of 2 tablets (group 1 = 1 donepezil 5 mg and 1 placebo, group 2 = 2 × 5 mg donepezil, and group 3 = 2 placebo tablets) At the end of the double-blind study all patients began a 3 week, single-blind washout period with placebo Other interventions used: use of any concomitant medications that could affect functioning of the central nervous system or interfere with efficacy assessments were prohibited. This included any	Number of participants: 468 patients were randomised: group 1) 157, group 2) 158, group 3) 153 Sample attrition/dropout: 56 patients withdrew from the trial (12%): Group 1: 16 (10%): adverse events 7 (4%); patient or investigator request 4 (3%), protocol violation 3 (2%), other 2 (1%) Group 2: 29 (18%): adverse events 14 (9%), Serious adverse events 2 (1%), patient or investigator request 6 (4%), protocol violation 4 (3%), other 3 (2%). Group 3 11 (7%): adverse events 2 (1%), serious adverse events 1 (1%), patient or investigator request 3 (2%), medication non-compliance 1 (1%), protocol violation 2 (1%), other 2 (1%). Adverse events and serious adverse events were not necessarily treatment emergent Sample crossovers: none reported Inclusion/exclusion criteria for study entry: males and females of any race aged ≥ 50 years. Diagnosis of probable AD consistent with the NINCD–ADRDA criteria and the DSM revised III edition categories 290.00 or 290.10. Those with mild to moderately severe disease as defined by MMSE scores of 10–26 and screening and baseline Clinical Dementia Rating (CDR) scores of 1 or 2. All underwent CT or MRI within 6 months of entry. Required to be ambulatory, or ambulatory when aided by either a walker or cane, and to have sufficient vision and hearing to enable them to comply with the study procedures Excluded if any of the following major medical illnesses: Type 1 diabetes; obstructive pulmonary disease or asthma; haematologic or oncologic disorders in the previous 2 years; vitamin B_{12} or folate deficiency. Also excluded if clinically significant active gastrointestinal, renal, hepatic, endocrine or cardiovascular system disease that was not well controlled by diet, pharmacological treatment or other therapeutic intervention. Those with evidence of other psychiatric or neurological disorders (e.g. stroke, schizophrenia or Parkinson's disease), and those with a Hachinski ischaemia score of 5 or more or known hypersensitivity to cholinesterase inhibitors were also excluded Characteristics of participants: states that no patient had AD that was complicated by delusions, delirium or depression, and none had a known or suspected history of alcoholism or drug abuse, but doesn't state whether these were	Primary outcomes: ADAS-cog, CIBIC-plus, Secondary outcomes: MMSE, the Sum of the Boxes of the Clinical Dementia Rating Scale (CDR-SB) and QoL. Adverse events Plasma concentrations of donepezil, clinical laboratory evaluations, blood pressure and ECG (data not extracted as per protocol). Also assessed compliance Methods of assessing outcomes: ADAS-cog is sensitive and reliable scale. CIBIC-plus is not a specific test instrument, and a variety of formats exist. The format chosen was a slightly modified version of the Alzheimer's Disease Cooperative Study – Clinical Global Impression of Change. Interviews of patient and caregiver were conducted by a clinician who was blinded from knowledge of other aspects of the study, including other test procedures, clinical laboratory values and adverse event reports. The MMSE is conducted by a trained clinician or psychometrician who evaluates the cognitive state of the patient. The CDR-SB is conducted as a consensus assessment by each patient's team, except the CIBIC-plus interviewer, and was based on information from all procedures conducted during a clinic visit. The QoL test was conducted through patient interviews by a nurse evaluator or another clinician. Although this hadn't been validated, it was selected because no QoL instrument has been validated in this patient population. Adverse events elicited at each visit by questioning the

continued

Reference and design	Intervention	Participants	Outcome measures
	anticholinergic, cholinomimetic, anticonvulsant, antidepressant, antipsychotic, antianxiety, or stimulating agents, as well as anti-Parkinson and certain antihypertensive agents. Occasional use of other medications, such as hypnotics and cold preparations (prescription and over-the-counter sympathomimetic amines and antihistamines) was allowed but not within 48–72 hours of a clinic visit. Approximately 90% of patients received allowable concomitant medication during the study	inclusion/exclusion criteria or by chance Gender, M/F, *n* (%): group 1 49 (31)/108 (69); group 2 62(39)/96 (61); group 3 60 (39)/93 (61). Age in years, mean ± SEM, (range): Group 1: 73.8 ± 0.67 (50–94); Group 2: 73.4 ± 0.65 (50–92); Group 3: 74.0 ± 0.65 (52–93). Weight in kg, mean ± SEM, (range): Group 1: 65.72 ± 0.98 (40.9–99.5); Group 2: 67.8 ± 1.13 (35.5–105.2); Group 3: 66.05 ± 1.01 (43.6–100.5) Race (White/Black/Other). *n* (%): Group 1: 149(95)/6(4)/2(1); Group 2: 152(96)/1(1)/5(3); Group 3: 147 (96)/6(4)/0 CDR, 0.5*/1.0/2.0, *n* (%): Group 1: 1(1)/121(77)/35(22); Group 2: 3(2)/120(76)/35(22); Group 3: 2(1)/121(79)/30(20). * These patients represented protocol violations and were subsequently discontinued from the study. MMSE, mean ± SEM (range): Group 1: 19.39 ± 0.39 (10–28); Group 2: 19.35 ± 0.40 (8–28); Group 3: 19.80 ± 0.35 (10–26). Only 8 patients had been previously treated with other cholinesterase inhibitors, 5 who had been enrolled in other investigative clinical trials	patient and caregiver, through direct observation by the treatment team Efficacy and safety assessment were undertaken at 3 week intervals throughout the trial. Compliance measured by counting returned tablets. Patients considered compliant when 80% or more was taken. Compliance was used as one of the determinants of the evaluable patient population (no data presented) Length of follow-up: 12 weeks

Results

Outcomes (states ITT using last observation carried forward, but some patients did not have an assessment)	**5 mg Donepezil (*n* = 156)**	**10 mg Donepezil (*n* = 155)**	**Placebo (*n* = 150)**	***p*-Value versus placebo (95% CI)**
ADAS-cog				
Mean ±SEM baseline (range)	26.4 ± 0.92 (5.7, 53.3)	26.4 ± 0.89 (4.7, 56.7)	25.3 ± 0.87 (6.0–51.3)	
Least-squares mean ± SEM change at endpoint	–2.1 ± 0.43*	–2.7 ± 0.43**	0.4 ± 0.43	*$p < 0.001$ (-3.59, –1.29) **$p < 0.001$ (–4.22, –1.92)
Least-squares mean ± SEM change at 15 weeks[a]	–0.7 ± 0.47*	–1.6 ± 0.49**	1.5 ± 0.47	*$p = 0.001$ **$p < 0.001$
CIBIC-plus	*n* = 153	*n* = 152	*n* = 150	
Mean ± SEM at endpoint	3.9 ± 0.08*	3.8 ± 0.08**	4.2 ± 0.07	*$p = 0.03$ (–0.50, –0.08) **$p = 0.08$ (–0.55, –0.13)
Mean ± SEM at 15 weeks[a]	4.0 ± 0.09	4.1 ± 0.09	4.2 ± 0.08	

[a]After 3 weeks of single-blind washout period

Comments: ADAS-cog consists of 11 items that evaluate selected aspects of memory, orientation, attention, language, reasoning and praxis. Scores range from 0 (no impairment) to 70 (very severe impairment). To reduce the potential for practice or carryover effects at subsequent visits, different word lists were used

continued

CIBIC-plus assesses patient function in 4 areas, general, cognitive, behavioural, and activities of daily living through examination of 15 separate domains. Disease severity is rated at baseline (CIBIS-plus). Using the baseline interview as the sole source for comparison, patients are reexamined at subsequent visits to determine whether their conditions had changed. The changes from baseline at subsequent visits (CIBIC-plus) is scored by the same interviewer using a 7-point Likert-type scale, in which 1 represents markedly improved, 4 no change, and 7 markedly worse.

The MMSE evaluates the cognitive state of the patient, including aspects of memory, orientation, language, and praxis.

The CDR-SB is a global scale that assesses 6 domains of patient function (memory, orientation, judgement and problem solving, community affairs, home and hobbies, and personal care).

The QOL assessment was a 7 item patient-rated scale that evaluated patients' perceptions of the well-being in terms of relationships, eating and sleeping, and social and leisure activities. The items were scored by marking on an analogue scale between 2 anchor points: the extremes were 0 (worst quality) and 50 (best quality).

Least-squares mean = adjusted for baseline severity.

In general the magnitude of improvement in mean change in ADAS-cog scores for the 10 mg group appears to be greater than the 5 mg group, but these did not reach statistical significance at endpoint ($p = 0.28$) by ANCOVA, although the study wasn't powered to detect such a difference.

ADAS-cog change from baseline during 12 weeks (estimated from figure)	**5 mg Donepezil (*n* = 156)**	**10 mg Donepezil (*n* = 155)**	**Placebo (*n* = 150)**	
3 weeks	–1.8 ($p = 0.015$)	–2.9 ($p < 0.001$)	–0.6	*p*-Values shown = versus placebo
6 weeks	–1.5 ($p = 0.011$)	–2.85 ($p < 0.001$)	–0.1	
9 weeks	–2.1 ($p = 0.005$)	–2.5 ($p < 0.001$)	–0.5	
CIBIC-plus during 12 weeks (estimated from figure)	***n* = 153**	***n* = 152**	***n* = 150**	
3 weeks	3.98	3.88	3.98	*p*-Values shown = versus placebo
6 weeks	3.84	3.93	3.98	
9 weeks	3.72 ($p = 0.45$)	3.88	4.07	

Comments: The percentage demonstrating clinical improvement at endpoint (score 1, 2, 3) were: 18% in the placebo group, 32% in the 5 mg donepezil group and 38% in the donepezil 10 mg group.

MMSE	**5 mg Donepezil (*n* = 156)**	**10 mg Donepezil (*n* = 156)**	**Placebo (*n* = 150)**	***p*-Value versus placebo (95% CI)**
Mean ± SEM baseline (range)	19.4 ± 0.39 (10–28)	19.3 ± 0.40 (8–28)	19.8 ± 0.35 (10–26)	
Least-squares mean ± SEM change at endpoint	1.0 ± 0.25*	1.3 ± 0.24**	0.04 ± 0.25	*$p < 0.004$ (0.33, 1.65) **$p < 0.001$ (0.65, 1.97)
Least-squares mean ± SEM change at 15 weeks[a]	0.7 ± 0.27	0.8 ± 0.28	–0.03 ± 0.27	
CDR-SB	***n* = 156**	***n* = 154**	***n* = 150**	***p*-Value versus placebo (95% CI)**
Mean ± SEM baseline	6.85 ± 0.18	7.18 ± 0.20	6.81 ± 0.18	
Least-squares mean ± SEM change at endpoint	–0.10 ± 0.11*	–0.31 ± 0.11	–0.14 ± 0.11	*$p = 0.32$ (–0.25, 0.33)
Adjusted mean ± SEM change at 15 weeks[a]	0.03 ± 0.13	–0.27 ± 0.13	0.03 ± 0.13	
QOL	***n* = 155**	***n* = 156**	***n* = 150**	***p*-Value versus placebo (95% CI)**
Mean ± SEM baseline	292.3 ± 3.6	283.5 ± 3.5	289.4 ± 3.4	
Least-squares mean ± SEM change at endpoint	5.7 ± 2.7*	–4.3 ± 2.7**	4.0 ± 2.7	*$p = 0.65$ (–5.58, 8.92) **$p = 0.02$ (–15.55, –1.07)
Least-squares mean ± SEM change at 15 weeks[a]	2.0 ± 2.8	–3.9 ± 3.0	5.6 ± 2.9	

[a]After 3 weeks of single-blind washout period

continued

MMSE least-squares mean change from baseline during 12 weeks (estimated from figure)	**5 mg Donepezil (*n* = 156)**	**10 mg Donepezil (*n* = 156)**	**Placebo (*n* = 150)**	
3 weeks	0.6 ($p = 0.03$)	1.2 ($p < 0.001$)	0.05	*p*-Values shown = versus placebo
6 weeks	0.8	1.4 ($p = 0.03$)	0.6	
9 weeks	1.1	1.2 ($p = 0.06$)	0.45	
CDR-SB least-squares mean change from baseline during 12 weeks (estimated from figure)	***n* = 156**	***n* = 154**	***n* = 150**	
3 weeks	–0.01	–0.24		Overall treatment effect
6 weeks	0.09	–0.28		10 mg group $p = 0.008$
9 weeks	–0.03	–0.29		
Adverse effects (*n* (%) with TESS)	**5 mg Donepezil, *n* = 157**	**10 mg Donepezil, *n* = 158**	**Placebo, *n* = 153**	***p*-Value comparing 3 groups**
No with ≥1 TESS	106 (68)	124 (78)	106 (69)	
Nausea	11 (7)	34 (22)	12 (8)	$p < 0.001$‡
Insomnia	13 (8)	28 (18)	8 (5)	$p = 0.001$‡
Diarrhoea	10 (6)	21 (13)	4 (3)	$p = 0.001$‡
Pain	14 (9)	21 (13)	11 (7)	$p = 0.20$
Headache	21 (13)	19 (12)	13 (8)	$p = 0.37$
Dizziness	14 (9)	14 (9)	10 (7)	$p = 0.69$
Muscle cramp	9 (6)	12 (8)	6 (4)	$p = 0.37$
Fatigue	5 (3)	12 (8)	8 (5)	$p = 0.22$
Accident	9 (6)	10 (6)	11 (7)	$p = 0.87$
Agitation	7 (4)	10 (6)	11 (7)	$p = 0.59$
Vomiting	5 (3)	10 (6)	7 (5)	$p = 0.41$
Anorexia	6 (4)	10 (6)	4 (3)	
Weight loss	3 (2)	8 (5)	3 (2)	
Common cold	8 (5)	7 (4)	10 (7)	$p = 0.69$
Abdominal disturbance	9 (6)	6 (4)	6 (4)	
Urinary tract infection	10 (6)	6 (4)	20 (13)	$p = 0.009$†
Stomach upset	8 (5)	5 (3)	1 (1)	
Rhinitis	8 (5)	5 (3)	6 (4)	
Upper respiratory tract infection	8 (5)	5 (3)	6 (4)	
Oedema in extremities	1 (1)	4 (3)	8 (5)	
Cough	2 (1)	3 (2)	8 (5)	

† More frequent with placebo, ‡ more frequent with donepezil

Comments: A high number of adverse events were reported for both the drug-treated and placebo groups. The incidence of treatment emergent signs and symptoms (TESS) for both doses of donepezil (68% for 5 mg group, 78% for 10 mg group) were comparable with the incidences observed with placebo (69%). In the majority of cases (92%) TESS were judged to be mild.

Seven patients treated with placebo and 6 in each of the donepezil groups suffered serious adverse events; three patients had events that were considered possibly related to treatment with donepezil. These included stomach ulcer with haemorrhage (5 mg group); syncope and transient ischaemic event (5 mg group); and nausea, aphasia, tremor and diaphoresis (10 mg group). One patient in the placebo group died of renal failure.

The incidence of adverse events was low overall, but higher in the group with 10 mg/day donepezil. The frequency was similar in the 5 mg/day group and the placebo group. The most common adverse events leading to discontinuation were nausea and diarrhoea, although these were rated as mild in general, and in most cases did not lead to discontinuation. In the 10 mg/day group 3.8% and 2% withdrew because of nausea and diarrhoea respectively.

continued

Methodological comments

- Allocation to treatment groups: Uses term randomised, but no methods reported.
- Blinding: States double-blind.
- Comparability of treatment groups: States that the 3 treatment groups were found to be comparable with respect to all demographic characteristics; no statistics presented.
- Method of data analysis: States that the primary analyses of efficacy and safety were performed on an intention-to-treat (ITT) population. For the safety analysis this included all patients who were randomised to receive treatment, while the analysis of efficacy (that requires change from baseline calculation) included all patients who had at least 1 postbaseline evaluation while undergoing treatment (therefore doesn't meet criteria for ITT). The primary analysis was conducted on the endpoint data set. Endpoint was week 12. For those not completing the study their last observation was carried forward and used as the end point value. Secondary analyses were also undertaken in the fully evaluable population to confirm the conclusions of the ITT analysis. Fully evaluable patients were those who completed the 12-week period of double-blind treatment and who had at least 80% medication compliance at the week 12 visit and at a minimum of 2 other visits during the trial. For continuous variables (ADAS-cog, MMSE, CDR-SB, and QoL) a general linear model was used to construct analysis of covariance models to compare the treatment groups with respect to changes from baseline in efficacy variables. After confirming the assumptions underlying ANCOVA, the reduced model contained effect for baseline scores (covariate), treatment effect and centre effect. Type III sums of squares were used to determine statistical significance among the 3 treatment groups. In cases where differences existed, pairwise comparisons of the groups were undertaken using Fisher 2-tailed least significant difference procedure. The CIBIC-plus was analysed using the Cochran–Mantel–Haenszel test, with RIDITS as the score option, including adjustment for centre. The analysis of adverse events was confined to treatment-emergent signs and symptoms (TESS) that began during or after administration of the first dose of study medication, or became more severe during treatment, compared by Fisher exact test.
  All *p*-values of 0.05 or less were considered statistically significant.
- Sample size/power calculation: Planned a study population of 150 patients based on a review of clinical studies of other cholinesterase inhibitors and the results of a previous phase 2 study with donepezil. The sample size was intended to provide 80% power to detect a 0.27-point difference in the mean CIBIC plus scores for donepezil treatment groups compared with placebo at the 5% significance level and assuming a patient completion rate of 80%. It was assumed that the dosages of 5 mg/day and 10 mg/day of donepezil would have equal efficacy, therefore the study was not powered to detect a difference between the active treatments but only between placebo and each active treatment group (assumption based on previous study results, and a review of studies).
- Attrition/dropout: Reports numbers and reasons (see above).

General comments

- Generalisability: Those with probable Alzheimer's disease, mild to moderate on MMSE.
- Outcome measures: Most psychological variables were reliable and valid. QoL not tested for reliability or validity.
- Inter-centre variability: not reported.
- Conflict of interests: Funded by pharmaceutical company.

Quality criteria Rogers *et al.*[52]

1. Was the assignment to the treatment groups really random?	Unknown
2. Was the treatment allocation concealed?	Unknown
3. Were the groups similar at baseline in terms of prognostic factors?	Reported
4. Were the eligibility criteria specified?	Adequate
5. Were outcome assessors blinded to the treatment allocation?	Adequate
6. Was the care provider blinded?	Partial
7. Was the patient blinded?	Partial
8. Were the point estimates and measure of variability presented for the primary outcome measure?	Adequate
9. Did the analyses include an intention-to-treat analysis?	Inadequate
10. Were withdrawals and dropouts completely described?	Adequate

Reference and design	Intervention	Participants	Outcome measures
Author: Seltzer *et al*.[55] Year: Country: International Study design: RCT Number of centres: 17 Funding: Eisai Inc and Pfizer Inc	**[Commercial/academic confidential information removed]**		

[Commercial/academic confidential information removed]

1. Was the assignment to the treatment groups really random?	
2. Was the treatment allocation concealed?	
3. Were the groups similar at baseline in terms of prognostic factors?	
4. Were the eligibility criteria specified?	
5. Were outcome assessors blinded to the treatment allocation?	
6. Was the care provider blinded?	
7. Was the patient blinded?	
8. Were the point estimates and measure of variability presented for the primary outcome measure?	
9. Did the analyses include an intention-to-treat analysis?	
10. Were withdrawals and dropouts completely described?	

Reference and design	Intervention	Participants	Outcome measures
Author: Winblad *et al*.[47] (and Wimo *et al.*, 2003[56]) Year: 2001 Country: Europe Study design: RCT, multicentre Number of centres: 28 centres in 5 countries Denmark Finland Norway Sweden The Netherlands Funding: Pfizer Pharmaceuticals Group, Pfizer, Inc.	Treatment arms: Two arms (1) donepezil 5 mg/day for 28 days and then 10 mg/day (as per clinician's judgement) (2) placebo Treatment continued for 1 year. If required, a dose reduction back to the 5 mg/day level was permitted. NB. results indicate that placebo could also be increased to 10 mg/day and reduced if necessary – presumably indicating blinding was maintained Over the course of the trial the dose of medication was increased from 5 to 10 mg/day in 91.5% of donepezil-treated patients and 97.2% of placebo-treated patients. The dose was decreased back from 10 to 5 mg/day in 9.9% of donepezil patients and 4.2% of placebo-treated patients. In 1.4% of the donepezil group the dose was increased again to 10 mg/day. Median duration of treatment was 361 days for both groups Other interventions used: The use of selective serotonin reuptake inhibitors,	Number of participants: 321 patients were screened (35 failures due to not meeting the entrance criteria $n = 27$; other reasons $n = 7$ and withdrawal of consent $n = 1$) and 286 patients were randomised. Donepezil $n = 142$ Placebo $n = 144$ Sample attrition/dropout: There were 94 withdrawals, 47 from the donepezil group (33.1%) and 47 from the placebo group (32.6%) Sample crossovers: None Inclusion/exclusion criteria for study entry: Men and women (who were at least 2 years postmenopausal or surgically sterile) of any race between 40 and 90 years of age who fulfilled all the inclusion and none of the exclusion criteria were accepted for enrolment Inclusion criteria Diagnosis of AD consistent with DSM-IV and NINCDS-ADRDA criteria for possible or probable AD. MMSE score of 10–26 inclusive. CT or MRI scans performed at screening if not performed within the last 12 months. Patients otherwise healthy, ambulatory or ambulatory with walker/cane etc. Vision and hearing sufficient to comply with testing procedures. Laboratory test values within normal limits or considered clinically insignificant by the investigator. All patients had to have a reliable caregiver Exclusion criteria Evidence of clinically significant and unstable active gastrointestinal, renal, hepatic, endocrine or cardiovascular system disease. Primary neurologic or psychiatric disease other than AD (notably DSM-IV-defined depression or vascular dementia), newly treated hypothyroidism, or a known or suspected history (within the past 10 years) of alcoholism or drug abuse. Evidence of neoplasm, insulin-dependent diabetes or diabetes not stabilised by diet or oral hypoglycaemic agents, obstructive pulmonary disease or asthma, recent (<2 years) haematologic/oncologic disorders, pernicious anaemia, or vitamin B_{12} or folate deficiency as evidenced by blood concentrations below the lower normal limit. Patients with a known hypersensitivity to ChE inhibitors, as well as those treated with cholinomimetics, including tacrine, within 30 days of screening were excluded	Primary outcomes: Gottfries-Bråne-Steen (GBS) scale. This comprehensive global assessment for rating dementia symptoms is based on a semistructured interview by the clinician with the patient and caregiver. There are four domains encompassing 27 items: GBS-I: intellectual impairment, contains 12 items including orientation, memory and concentration. GBS/ADL: motoric function contains six items assessing primarily self-care or basic activities of daily living. GBS-E: emotional reaction/function, 3 items. GBS-S: Dementia symptoms, six items representing pathologic aspects of behaviour Secondary outcomes: MMSE, ADL (Progressive Deterioration Scale [PDS]), NPI, GDS Apolipoprotein E genotyping (data not extracted as per protocol) Safety From Wimo ref: Instrumental Activities of Daily Living (IADL) scale, Physical Self-Maintenance Scale (PSMS). No details provided Methods of assessing outcomes: The primary outcome, GBS, was assessed by the clinician using a semistructured interview with the patient and caregiver. No further specific details are given Length of follow-up: 52 weeks

continued

Reference and design	Intervention	Participants	Outcome measures
	neuroleptics in small daily doses, and short-acting benzodiazepines was permitted provided that they were given in stable doses for at least 2 months before entering the study	Concomitant medications: Medications with major anticholinergic effects, such as high doses of neuroleptics, tricyclic antidepressants, and medications for PD, were not permitted. Prescription or over-the-counter sympathomimetic amines and antihistamines were to be stopped temporarily for 48 hours before each clinic visit. This study began before the potential benefits of compounds such as vitamin E, gingko biloba, and memantine to patients with AD had been reported therefore there were no specific requirements concerning the use of these compounds Characteristics of participants: Gp1 = donepezil (n = 142) Gp2 = placebo (n = 144) Mean age, Years ± SD (range) Gp1 72.1 ± 8.6 (49–86); Gp2 72.9 ± 8.0 (51–88) Number female patients (%) Gp1 99 (69.7); Gp2 85 (59.0) Race, white (%) Gp1 142 (100); Gp2 144 (100) APOE4 positive (homo- or heterozygous) Gp1 98; Gp2 98. Female % Gp1 72/142 (50.7); Gp2 59/143 (41.3) Mean baseline MMSE score ± SD (range) Gp1 19.37 ± 4.37 (10–26); Gp2 19.26 ± 4.54 (10–26) Presence of at least one abnormal medical history finding or comorbid illness Gp1 117/142 (82%); Gp2 127/144 (88%) Concomitant medication use Gp1 129/142 (91%); Gp2 132/144 (92%) Antihypertensive drug use (all drugs) Gp1 12/142; Gp2 24/144 (there were no differences in the use of hypertensives with known centrally acting effects) Vitamin E use Gp1 2/142 Gp2 3/144 Use of psychotropic medications (e.g. hypnotics, sedatives, anxiolytics) Gp1 32.4% Gp2 27.8% The following baseline scores are all presented as Mean ± SD GBS score: Gp1 29.51 ± 17.33; Gp2 29.77 ± 17.84 GBS-I score: Gp1 18.42 ± 10.00; Gp2 18.21 ± 10.08 PDS score: Gp1 52.77 ± 20.58; Gp2 52.93 ± 20.45 GDS score: Gp1 4.15 ± 0.83; Gp2 4.16 ± 0.90 NPI score: Gp1 13.05 ± 13.76; Gp2 11.78 ± 12.23	Efficacy measures were performed at screening, baseline, and at weeks 4, 12, 24, 36 and 52. Safety and compliance were also assessed at these time points and at week 6

continued

Results

Outcomes	**Donepezil (*n* = 142)**	**Placebo (*n* = 144)**	***p*-Value**
Mean overall rate of compliance	94.6%	94.9%	
LS Mean change from baseline in GBS total scores ± SE. *italics* = estimated from graph			
Week 4	*–0.5 ± 0.5* (n = *136*)	*0.3 ± 0.9* (n = *142*)	
Week 12	*1.6 ± 1.0* (n = *129*)	*2.8 ± 1.2* (n = *129*)	
Week 24	*1.7 ± 1.2* (n = *122*)	*5.0 ± 1.4* (n = *121*)	$p = 0.046$
Week 36	*3.8 ± 1.7* (n = *105*)	*9.2 ± 1.7* (n = *105*)	$p = 0.012$
Week 52	*7.3 ± 2.1* (n = 93)	*13.5 ± 2.1* (n = 97)	$p = 0.014$
Endpoint (LOCF)	*8.0 ± 1.4* (n = *138*)	*11.5 ± 1.6* (n = *144*)	$p = 0.054$
% Improved from baseline (categorical analysis of GBS total scores)	Week 12 40.3% Week 24 38.5% Week 52 31.2%	Week 12 31.0% Week 24 32.2% Week 52 21.6%	Don't know whether these values ITT – would think not
GBS-I score LS Mean change from baseline ± SE			Probably not ITT
Week 52	3.6 ± 1.1	7.3 ± 1.1	$p = 0.004$
Comments: GBS-I score (this domain contains 12 items) treatment differences in favour of donepezil were also observed at weeks 24 ($p = 0.049$), 36 ($p = 0.003$) and at the endpoint ($p = 0.012$). Although the decline on the GBS/ADL, GBS-E and GBS-S subtotals was smaller in donepezil- compared with placebo-treated patients at weeks 24, 36, and 52, there were no significant differences between the treatment groups at week 52. This might have been due to the minimal impairment in these domains at baseline. GBS Scale measures: A 7-point scoring system from 0 to 6 is used for each of the 27 items (from the 4 domains) of the scale, giving a score range of 0 to 162. *An increase in score represents clinical deterioration*. This scale has been demonstrated to be highly reliable across a number of countries and languages.			
LS Mean change from baseline in MMSE scores ± SE. *italics* = estimated from graph			
Week 12	*0.69 ± 0.23* (n = *127*)	*–0.11 ± 0.29* (n = *128*)	$p = 0.053$
Week 24	*0.40 ± 0.34* (n = *121*)	*–1.09 ± 0.34* (n = *120*)	$p < 0.001$
Week 36	*0.00 ± 0.40* (n = *104*)	*–1.15 ± 0.40* (n = *105*)	$p = 0.019$
Week 52	*–0.34 ± 0.52* (n = *91*)	*2.23 ± 0.46* (n = *98*)	$p = 0.001$
Endpoint (LOCF)	*–0.46 ± 0.34* (n = *135*)	*–2.18 ± 0.29* (n = *137*)	$p < 0.001$
Comments: MMSE change score – positive score is clinical improvement			
Activities of daily living LS Mean change from baseline. *italics* = estimated from graph			
Overall PDS score at week 52	*–10.8*	*–15.3*	$p < 0.05$
Finance	*–10.8*	*–13.6*	
Social interaction	*–8.5*	*–13.6*	
Hobbies/leisure	*–12.2*	*–11.9*	
Safe driving	*7.5* (*n* = 47)	*–12.2* (*n* = 55)	
Spatial orientation	*–12.9*	*–17.5*	
Telephone	*–10.8*	*–21.4*	$p < 0.01$
Time	*–16.3*	*–23.7*	
Task performance	*–11.5*	*–14.9*	
Memory	*3.1*	*–6.1*	$p < 0.01$
Self care	*–9.5*	*–16.6*	$p < 0.05$
Comments: A positive score indicates clinical improvement. Less deterioration in ADL was observed for donepezil- compared with placebo-treated patients as assessed by the LS mean change in scores from baseline on the PDS at all postbaseline evaluations, which was significant at week 52 (this is the data shown above, not ITT) and the endpoint (data not reported in paper but would have been ITT). For the week-52 results shown above data were available for 89 to 93 donepezil- and 94 to 97 placebo-treated patients. For the safe driving item, because not all of the patients were drivers, data were available for 47 donepezil- and 55 placebo-treated patients.			

continued

LS Mean change from baseline in GDS scores ± SE *italics* = estimated from graph			
Week 12	*0.00 ± 0.04 (n = 128)*	*0.09 ± 0.03 (n = 130)*	
Week 24	*0.01 ± 0.06 (n = 122)*	*0.17 ± 0.06 (n = 121)*	$p = 0.026$
Week 36	*0.10 ± 0.08 (n = 105)*	*0.38 ± 0.08 (n = 105)*	$p = 0.004$
Week 52	*0.19 ± 0.08 (n = 93)*	*0.47 ± 0.08 (n = 98)*	$p = 0.011$
Endpoint (LOCF)	*0.25 ± 0.06 (n = 136)*	*0.44 ± 0.06 (n = 140)*	$p = 0.014$
% of patients demonstrating a postbaseline improvement in GDS score. *italics* = estimated from poor quality figure			
Week 12	*10% (n = 128)*	*6% (n = 130)*	
Week 24	*12% (n = 122)*	*7% (n = 121)*	
Week 36	*14% (n = 105)*	*7% (n = 105)*	
Week 52	*14% (n = 93)*	*5% (n = 98)*	$p = 0.047$

Comments: Patients in the donepezil group showed significant benefits over placebo at weeks 24, 36, 52 and at endpoint. Categorical analysis (i.e. as % improved) demonstrated that approximately twice as many donepezil- as placebo-treated patients showed postbaseline improvement at weeks 12, 24, and 52. Increasingly more patients treated with donepezil improved from baseline over time until week 52, according to the GDS, than those receiving placebo ($p = 0.047$).

NPI total scores	Differences in favour of donepezil were observed from week 12 onward for the LS mean change from baseline but these were not significant (despite use of psychotropic medications by a large proportion of patients in both groups). Note that the baseline scores reflect a patient population relatively unimpaired with respect to behavioural abnormalities.

IADL, % deteriorating (estimated from figure in Wimo 2003)	***n* = 135–136***	***n* = 138–40***	***p*-Value**
Telephone	31	39	Overall $p = 0.025$
Shopping	36	47	
Food preparation	25	38	
Housekeeping	36	41	
Laundry	32	25	
Transportation	41	35	
Medication	22	22	
Finances	27	28	

Comments: *the exact number of patients depended on the specific item of the IADL

From Wimo 2003: PSMS, % deteriorating reported in text (no data) that significantly fewer donepezil patients experienced an overall decline in ADL than placebo-treated patients at weeks 24 ($p = 0.011$) and 36 ($p = 0.032$). Data not reported for 52 weeks

Completed (%)/Withdrawn(%)	$n = 95$ (66.9) $n = 47$ (33.1)	$n = 97$ (67.4) $n = 47$ (32.6)
Reasons for withdrawals	5 (3.5%)	6 (4.2%)
study drug related	5 (3.5%)	3 (2.1%)
not related	4 (2.8%)	6 (4.2%)
Insufficient clinical response	4 (2.8%)	3 (2.1%)
Subject died	20 (14.1%)	18 (12.5%)
Withdrew consent	9 (6.3%)	11 (7.6%)
Other		
Discontinuations due to adverse events	7%	6.3%

Comments: The largest number of discontinuations in both treatment groups occurred between weeks 24 and 36 (34.0% donepezil vs 46.8% placebo), which coincided with the launch of the study drug in three of the participating countries (Sweden, Denmark, Finland). Of the 94 patients who discontinued prematurely, 52 returned for the week 52 follow-up visit. At this visit, 21 donepezil- and 24 placebo-treated patients indicated that their withdrawal was due to the commercial availability of donepezil, and most continued on donepezil treatment after withdrawing from the trial. No significant difference in the time to discontinuation between the treatment groups was demonstrated for discontinuations due to all causes, lack of efficacy, or safety reasons as assessed by Kaplan–Meier survival analyses.

continued

No. of patients with treatment-emergent adverse events that occurred in ≥5% of patients in either treatment group	116 (81.7)	109 (75.7)
With adverse events (%)		
Nausea	16 (11.3)	13 (9.0)
Depression	16 (11.3)	11 (7.6)
Anxiety	15 (10.6)	8 (5.6)
Insomnia	14 (9.9)	10 (6.9)
Asthenia	11 (7.7)	5 (3.5)
Headache	11 (7.7)	9 (6.3)
Vertigo	11 (7.7)	3 (2.1)
Diarrhoea	10 (7.0)	10 (6.9)
Syncope	9 (6.3)	4 (2.8)
Bone fracture (accidental)	8 (5.6)	5 (3.5)
Dizziness	9 (6.3)	6 (4.2)
Urinary tract infection	8 (5.6)	10 (6.9)
Constipation	6 (4.2)	9 (6.3)
Confusion	4 (2.8)	9 (6.3)
Hostility	4 (2.8)	8 (5.6)
Abdominal pain	3 (2.1)	8 (5.6)
Of the AEs, most were:		
Mild adverse events	44 (31)	59 (41)
Moderate adverse events	45 (31.7)	38 (26.4)

Comments: Vertigo, asthenia, and syncope occurred at least at twice the rate in the donepezil group as in the placebo group but none was considered due to study drug and all patients continued taking the drug.

No. of patients with treatment-emergent serious adverse events that occurred in at least 2 patients in either treatment group	35 (24.6)	20 (13.9)
With serious adverse events		
Bone fracture (accidental)	6 (4.2)	3 (2.1)
Syncope	3 (2.1)	1 (0.7)
Headache	3 (2.1)	0
Myocardial infarct	2 (1.4)	1 (0.7)
Nausea	2 (1.4)	1 (0.7)
Urinary tract infection	2 (1.4)	1 (0.7)
Accidental injury	2 (1.4)	0
Pneumonia	2 (1.4)	0
Confusion	1 (0.7)	2 (1.4)
Procedure (medical/surgical/health)	0	2 (1.4)

Comments: Only one patient in the donepezil group reported a serious AE (moderate nausea) that was considered to be related to the study drug, and this patient continued taking the drug.

Methodological comments

- Allocation to treatment groups: Randomisation was performed using a computer-generated randomisation list produced by Pfizer, Inc (New York).
- Blinding: The study is described as double-blinded but other than that no details are given as to how this was achieved and maintained.
- Comparability of treatment groups: Groups appear similar at baseline.
- Method of data analysis: Efficacy analyses at the end of the study were performed on the ITT populations, defined as all patients who were randomised to treatment, received at least one dose of study medication, and who provided data at baseline and at least one postbaseline efficacy assessment. In the ITT population analyses of both observed cases (OC) at each scheduled visit (week 4, 12, 36 and 52) and last observation carried forward (LOCF) at week 52 were conducted; only the week 52 LOCF ITT results are presented in the paper. Week 52 LOCF using the GBS total score was defined as the primary endpoint evaluation for each patient. Non-ITT results are also presented for the intermediate results at weeks 4, 12, 24, 36 and also week 52; in most cases the number of patients contributing to the outcomes at these time points is given. The analysis of safety was performed on the population that included all patients who received at least one dose of study medication and who provided any postbaseline follow-up data. Most data are presented as means and SE. For all efficacy variables, an ANCOVA model was used for estimating and testing treatment effects. All statistical tests were two-sided and a p-value of <0.05 was considered significant. In addition PDS subdomain analysis was explored.

continued

No *a priori* statistical tests were planned on safety data but group summaries were performed. Categorical analyses were used to calculate the percentage of patients in each treatment group showing postbaseline improvement for the GBS and GDS assessments at various time points. Treatment effects were tested using Fischer's exact test.

- Sample size/power calculation: A target sample size of approximately 150 patients per treatment group was determined using the primary efficacy variable (GBS) to achieve a power of 0.8 ($\alpha = 0.5$). This estimate was based on previous clinical experience with the GBS scale suggesting that a 10% to 15% change from baseline in total score is clinically relevant. Sample sizes fell slightly short of this target: donepezil ($n = 142$) placebo ($n = 144$). The GBS outcome was reported but for the ITT population at the study end significance was borderline ($p = 0.054$). This is not discussed further.
- Attrition/dropout: The study was completed by 66.9% of donepezil- and 67.4% of placebo-treated patients. Attrition/dropouts were well reported. Over the year-long course of the study about one-third of patients dropped out from each group. How this may have affected outcomes is not discussed.

General comments

- Generalisability: Patients with mild to moderate AD (at least 2 years postmenopausal or surgically sterile). The inclusion criteria used limited the study population to those who were generally healthy although a range of concomitant medications were permitted providing doses had been stable for the preceding 2 months. The results might not be generalisable to patients with co-existing disease such as diabetes, COPD or asthma, cancer, blood disorders, etc.
- Outcome measures: Outcome measures appear appropriate, measurements should have been well conducted as raters received training but whether they were blinded or not is unclear.
- Inter-centre variability: Not reported.
- Conflict of interests: Pfizer Pharmaceuticals Group, Pfizer, Inc.

Quality criteria for Winblad *et al.*[47]

1. Was the assignment to the treatment groups really random?	Adequate
2. Was the treatment allocation concealed?	Unknown
3. Were the groups similar at baseline in terms of prognostic factors?	Reported
4. Were the eligibility criteria specified?	Adequate
5. Were outcome assessors blinded to the treatment allocation?	Unknown
6. Was the care provider blinded?	Partial
7. Was the patient blinded?	Partial
8. Were the point estimates and measure of variability presented for the primary outcome measure?	Adequate
9. Did the analyses include an intention-to-treat analysis?	Inadequate
10. Were withdrawals and dropouts completely described?	Adequate

Appendix 8

Data extraction: rivastigmine RCTs

Reference and design	Intervention	Participants	Outcome measures
Author: Agid *et al*.[59] Year: 1998 Country: France Study design: RCT Number of centres: 54 (Austria, Belgium, Czechoslovakia, Denmark, Finland, France, Germany, Ireland, Norway, Sweden, Switzerland, UK). Funding: Novartis Pharma	Treatment arms: (1) Rivastigmine 4 mg/day (low dose group). The dose was titrated to 2–3 mg/day over 1 week and after was maintained at 4 mg/day for 10 weeks (2) Rivastigmine 6 mg/day (high dose group). The dose was titrated to 2–5 mg/day over 3 weeks and after was maintained at 6 mg/day for 10 weeks (3) Placebo Following a 2-week placebo washout phase, patients wishing to continue therapy entered an extension phase. Other interventions used: before entry, all medication with cognitive enhancing potential was withdrawn for a minimum of 3 weeks. Medications for non-cognitive aspects of AD, such as hypnotics, were permitted provided they were not long-acting agents. Drugs for other concomitant conditions were continued at the same doses as before the study	Number of participants: total $n = 402$; group 1) $n = 136$, group 2) $n = 133$, group 3) $n = 133$ Sample attrition/dropout: 119 (87.5%) in the low dose group; 113 (85.0%) in the high dose group; 125 (94.0%) in the placebo group completed the study A total of 386 patients (169 men and 217 women; age range 50–90 years) received at least one dose of rivastigmine with a baseline measurement and at least one post-baseline measurement of at least one variable Sample crossovers: not reported Inclusion/exclusion criteria for study entry: enrolled on the basis of a diagnosis of mild-to-moderate demential (DSM III), and probable Alzheimer's disease (NINCDS-ADRDA) Characteristics of participants: mean ages (±SD): low dose 68.62 ± 8.64 years; high dose 68.68 ± 7.85 years; placebo 70.80 ± 8.58 years	Primary outcomes: proportion with "successful outcome" on the Clinical Global Impression of Change scale (CGIC); Fuld Object-Memory Evaluation (FOME); digit Symbol Substitution test (DSST); Benton Visual Retention Test (BVRT); Trail making test (TMT); Mini-mental State Examination (MMSE); Nurses' Observation Scale for Geriatric Patients (NOSGER); assessment of individual daily activities by both patients and caregivers; adverse events (by questioning patients and caregivers). Safety evaluations (not data extracted as per protocol). Methods of assessing outcomes: not reported except for daily activities and adverse events (noted above) Length of follow-up: at 13 weeks

Results

Outcomes	Low dose ($n = 119$)	High dose ($n = 113$)	Placebo ($n = 125$)	*p*-Value
Global/cognition (13 weeks unless stated)	**Low dose ($n = 111$)**	**High dose ($n = 103$)**	**Placebo ($n = 117$)**	***p*-Value compared to placebo**
% with 'successful' CGIC	31.53	42.72*	29.91	$^{*}p = 0.05$
DSST				
Week 7	2.1 ± 5.8	2.0 ± 5.4*	0.1 ± 7.4	$^{*}p \leq 0.005$
Week 13	1.7 ± 5.1	2.8 ± 8.1†	0.5 ± 6.9	$^{\dagger}p \leq 0.05$
FOME (total storage)				
Week 7	2.2 ± 7.3‡	2.0 ± 6.6†	0.0 ± 6.2	$^{\dagger}p \leq 0.05$
Week 13	0.4 ± 6.2‡	0.7 ± 6.2†	–0.9 ± 5.5	$^{\ddagger}p \leq 0.01$

continued

FOME (total retrieval)				
Week 7	1.7 ± 5.3†	2.4 ± 4.8*	0.5 ± 4.6	*$p \leq 0.005$
Week 13	0.8 ± 4.6	1.1 ± 4.2†	0.1 ± 4.3	†$p \leq 0.05$
NOSGER				
memory	0.7 ± 2.8	–0.2 ± 2.4	0.0 ± 3.4	
IADL†	0.0 ± 3.3	–0.7 ± 3.5	–0.2 ± 3.3	
Self-care	0.2 ± 2.8	–0.1 ± 2.1	0.1 ± 2.7	
Mood	0.2 ± 2.8	0.1 ± 2.5	0.1 ± 3.1	
Social behaviour	–0.3 ± 3.1	–0.5 ± 3.0	0.0 ± 3.6	
Disturbing behaviour	0.2 ± 2.2	–0.5 ± 2.3	0.0 ± 2.1	
MMSE	0.0 ± 3.3	0.3 ± 3.1	–0.0 ± 2.6	
BVRT				
Week 7	0.4 ± 2.2	0.7 ± 2.7	0.1 ± 2.4	
Week 13	0.3 ± 2.6	0.2 ± 2.6	0.2 ± 2.7	
TMT				
Week 7	–4.3 ± 36.9	–5.4 ± 45.3	–0.6 ± 31.2	
Week 13	–1.6 ± 39.0	–7.3 ± 48.9	0.5 ± 28.7	

Comments: Scores CGIC: 1 = marked improvement, greatly improved daily living function, 2 = moderate improvement, some improvements in daily living function, 3 = minimally improved, but no consequences for daily living function, 4 = no change, 5 = minimally worse, but no consequences for daily living function, 6 = moderately worse but some deterioration in daily living function, 7 = much worse, marked deterioration in daily living function. Patients with scores of 1 or 2 were classified as "successful outcome" and patients with scores of 3 to 7 were classified as "failures".

† Instrumental Activities of Daily Living

Adverse effects (incidence)	**Low dose (*n* = 136)**	**High dose (*n* = 133)**	**Placebo (*n* = 133)**	**Total**
Nausea	23 (17)	41 (31)	8 (6)	72 (18)
Vomiting	13 (10)	24 (18)	4 (3)	41 (10)
Diarrhoea	9 (7)	16 (12)	2 (2)	27 (7)
Abdominal pain	8 (6)	9 (7)	7 (5)	24 (6)
Dizziness	8 (6)	26 (20)	9 (7)	43 (11)
Headache	6 (4)	17 (13)	8 (6)	31 (8)
Withdrawals, *n* (%)				
Total withdrawals	17 (13)	20 (15)	8 (6)	45 (11)
Due to adverse events	14 (10)	16 (12)	5 (4)	35 (9)
Other	3 (2)	4 (3)	3 (2)	10 (2)
Comments				

Methodological comments

- Allocation to treatment groups: Participants assigned a randomisation number in chronological order by the investigator in chronological order according to a list generated by Novartis Pharma.
- Blinding: Active medication and placebo had the same physical appearance, and the number of capsules for each dose was the same in all three groups. Incidence of adverse events may unblind participants so can't be sure of double blinding remained throughout.
- Comparability of treatment groups: Minimal baseline characteristics reported.
- Method of data analysis: Treatment effects on the CGIC were analysed by the Van Elteren test and on the outcome CGIC by Maentel–Haenszel. Comparison between pairs of treatment groups for change from baseline in psychometric test scores was done using a modified Mantel–Haenszel test. No intention-to-treat analysis.
- Sample size/power calculation: Not reported.

continued

- Attrition/dropout: Numbers and some reasons reported (but different numbers included in the psychological outcomes reported so some lost to follow-up).

General comments
- Generalisability: Minimal inclusion criteria and minimal baseline characteristics reported therefore difficult to assess generalisability. Patients were elderly with mild to moderate dementia.
- Outcome measures: No details of reliability or validity of psychological variables.
- Inter-centre variability: Not reported.
- Conflict of interests: Funded by pharmaceutical company.

Quality criteria for Agid *et al.*[59]

1. Was the assignment to the treatment groups really random?	Inadequate
2. Was the treatment allocation concealed?	Unknown
3. Were the groups similar at baseline in terms of prognostic factors?	Reported
4. Were the eligibility criteria specified?	Partial
5. Were outcome assessors blinded to the treatment allocation?	Unknown
6. Was the care provider blinded?	Adequate
7. Was the patient blinded?	Adequate
8. Were the point estimates and measure of variability presented for the primary outcome measure?	Adequate
9. Did the analyses include an intention-to-treat analysis?	Inadequate
10. Were withdrawals and dropouts completely described?	Inadequate

Reference and design	Intervention	Participants	Outcome measures
Author: Corey-Bloom *et al*.[57] Year: 1998 Country: USA Study design: RCT Number of centres: 22 Funding: supported by Novartis pharmaceutical corporation	Treatment arms: (1) low dose group 1–4 mg/day after initial 7 weeks (see below) (2) high dose 6–12 mg/day after initial 7 weeks (see below) (3) placebo (throughout the study patients received 2 capsules daily with food) All patients had an initial fixed dose-titration phase through to week 7, followed by a flexible phase during weeks 8–26 Fixed dose-titration phase: any occurrence of adverse events then the same dosage level could be continued for up to 2 weeks; if didn't abate a dose decrease was not permitted, and participant could be discontinued if necessary. By week 7 had to tolerate the minimum dose requirement within the assigned dosage range (low-dose group 1 mg/day; high dose group 6 mg/day) or they had to be discontinued Flexible-dose phase: doses were further increased within the assigned range, until maximum dose or tolerated dose was achieved. Dose decreases were permitted as long as maintained within the assigned range At study end: mean dose in group 1) was 3.5 mg and in group 2)	Number of participants: total group 925, 226 not eligible (randomisation 699). Group 1) $n = 233$; 2) $n = 231$; 3) $n = 235$ Sample attrition/dropout: lost to follow-up (at each assessment): group 1) at 12 weeks $n = 223$ (10), 18 weeks $n = 208$ (25), 26 weeks $n = 194$ (39) group 2) at 12 weeks $n = 169$ (62), 18 weeks $n = 157$ (74), 26 weeks $n = 145$ (86) group 3) at 12 weeks $n = 216$ (19), 18 weeks $n = 201$ (34), 26 weeks $n = 192$ (43) Withdrawals: group 1) $n = 34$: withdrawal consent 10, failure to return 1, adverse event 19, other 4 group 2) $n = 82$: withdrawal consent 9, failure to return 2, adverse event 66, death 1, other 4 group 3) $n = 39$: withdrawal consent 10, treatment failure 4, adverse event 17, concurrent illness 1, non-compliance 1, other 6 Sample crossovers: none Inclusion/exclusion criteria for study entry: between 45–89 years (of non childbearing potential); fulfilling criteria for dementia of Alzheimer type by DSM-IV and NINCDS-ADRDA whose impairment was mild-moderately severe on MMSE (10–26). A head CT or MRI consistent with AD within 12 months of inclusion also required. Each had a responsible caregiver, and along with the caregiver provided written, informed consent. Excluded those with severe and unstable medical illnesses Characteristics of participants (groups 1–3): Age years, mean (range): 1) 74.9 (45–89), 2) 73.8 (50–89), 3) 74.8 (45–89). (p = ns) Age (%) ≤ 65 years: 1) 23 (10), 2) 35 (15), 3) 25 (11); 66–75 years: 1) 89 (38), 2) 95 (41), 3) 93 (40); 76–85 years: 1) 112 (48), 2) 87 (38), 3) 106 (45); >85 years: 1) 9 (4), 14 (6), 11 (5) Sex (%) Male/Female: 1) 100 (43)/133 (57), 2) 75 (32)/156 (68), 3) 98 (42)/137 (58). ($p < 0.041$) Race (%) Caucasian/Black/Asian, Oriental/Other: 1) 221 (95)/9(4)/0/3(1); 2) 225 (97)/6(3)/0/0; 3) 222 (94)/10(4)/2(1)/1 (<1). (p = ns) Dementia duration: mean (range) in months: 1) 39.3 (3–138); 2) 38.4 (5–126); 3) 40.4 (6–180)	Primary outcomes: cognitive subscale of the Alzheimer's Disease Assessment scale (ADAS-cog) assessing memory, language, orientation, simple tasks; structured Clinician's Interview Based Impression of Change incorporating clinical and caregiver information (CIBIC-plus) assessing behaviour, cognition, activities of daily living (ADLs), information from patient and caregiver; progressive deterioration scale (PDS) assessing ADLs informed by the caregiver Secondary outcomes: Staging measures: Mini-Mental State Examination (MMSE) assessing staging and cognition by the patient; Global Deterioration Scale (GDS) assessing memory, self-care, staging and cognition, ADLs, clinician-based using information by patient and caregiver. Adverse events (coded by Sandoz Medical Technology Thesaurus). Also safety evaluations (physical, ECGs, vital signs, laboratory) but data not extracted as per protocol Outcomes assessed at baseline, weeks 12, 18, and 26 or early termination Methods of assessing outcomes: unclear whether all patients were assessed by the same clinician on psychological outcomes Length of follow-up: 26 weeks study

continued

Reference and design	Intervention	Participants	Outcome measures
	was 9.7 mg. 83% in group 1 were receiving maximum dose of 4 mg at 26 weeks, and 55% of group 2 patients were receiving maximum dose of 12 mg at 26 weeks Other interventions used: patients allowed to continue most medications for coexistent diseases; however anticholinergic drugs, acetylcholine precursor health food supplements, memory enhancers, insulin, and psychotropic drugs were not permitted other than the occasional use of chloral hydrate (500 mg/day) for agitation/insomnia	Severity of disease from DSM-IV (%): Mild/Moderate/Severe: 1) 96(41)/133(57)/4(2); 2) 90(39)/139(60)/2(1); 3) 105(45)/129(55)/1(<1) (p = ns) ADAS-cog: 1) 22.4; 2) 22.3, 3) 21.7 GDS: 1) 4.0; 2) 4.0; 3) 3.9 MMSE: 1) 19.5; 2) 19.62; 3) 20 (p = ns) Concurrent medical conditions reported for total population: cardiovascular (35%), gastrointestinal (26%), musculoskeletal (44%), nervous (22%) Medications for other conditions for total population: cardiovascular (43%), gastrointestinal (59%), nervous system (primarily analgesics) (56%)	

Results

Outcomes	Rivastigmine 1–4 mg (n = 233)	Rivastigmine 6–12 mg (n = 231)	Placebo (n = 234)	Treatment difference (group 2 vs placebo)
Cognition (ITT analysis) change scores (95% CI) at 26 weeks	**Rivastigmine 1–4 mg (n = 233)**	**Rivastigmine 6–12 mg (n = 231)**	**Placebo (n = 234)**	**Treatment difference (group 2 vs placebo)**
ADAS-cog (NB. data in table have reversed signs)	2.36 (3.13, 1.59)	0.31 (1.08, –0.46)	+4.09 (4.86, 3.32)	3.78 (2.69, 4.87), $p < 0.001$
CIBIC-plus	0.23 (0.07, 0.39)	0.20 (0.04, 0.36)	0.49 (0.33, 0.65)	–0.29 (–0.51, –0.07), $p < 0.01$
PDS	–5.19 (–6.52, –3.86)	–1.52 (–2.85, –0.19)	–4.90 (–6.22, –3.58)	3.38 (1.51, 5.25), $p < 0.001$
GDS	–0.16 (–0.25, –0.07)	–0.13 (–0.22, –0.04)	–0.32 (–0.41, –0.23)	0.19 (0.06, 0.32), $p < 0.03$

Comments. Analysis on group 1 versus group 3, or group 1 versus group 2 not reported.
ADAS-cog: scale from 0–70 with 70 = severe impairment.
CIBIC-plus: scale from 1–7 where 1, 2, 3 = improvement, 4 = no change, 5, 6, 7 = deterioration.
PDS: scale not reported.
MMSE: scale from 0–30 where 0 = severe impairment.
GDS: scale from 1–7 where 1 = no cognitive decline, 7 = severe cognitive decline.

continued

QoL (observed case analysis) change scores (95% CI) at 26 weeks	Rivastigmine 1–4 mg (n = 233)	Rivastigmine 6–12 mg (n = 231)	Placebo (n = 234)	Treatment difference (group 2 vs placebo)
ADAS-cog	–2.27 (–3.13, –1.41)	0.79 (–0.22, 1.90)	–4.15 (–5.00, –3.30)	4.94 (3.62, 6.26), $p < 0.001$
CIBIC-plus	0.16 (–0.01, 0.33)	0.13 (–0.07, 0.33)	0.48 (0.31, 0.65)	–0.35 (–0.61, –0.09), $p < 0.01$
PDS	–5.25 (–6.81, –3.69)	–1.15 (–2.96, 0.66)	–5.69 (–7.23, –4.15)	4.54 (2.16, 6.92), $p < 0.001$
GDS	–0.15 (–0.25, –0.05)	–0.14 (–0.25, –0.03)	–0.33 (–0.42, –0.24)	0.19 (0.05, 0.33), $p < 0.012$
MMSE (low dose data estimated from figure)	–0.34	0.30	–0.79	$p < 0.05$

Comments. ADAS-cog scores significantly different between group 1 and placebo, $p < 0.05$. Reports that a significantly higher percentage of treated patients demonstrated clinically meaningful improvement on the ADAS-cog, (≥ 4 points). At 26 weeks $^1/_4$ of high dose participants showed clinically meaningful improvement as compared to placebo (data not reported). 56% of high dose group showed improvement or no change from baseline at 26 weeks, whereas 27% of placebo didn't decline (no statistics reported). Of the 73% placebo participants who declined on scores, 44% declined ≥ 4 points, and 30% by ≥ 7 points. Of high dose participants, only 21% declined ≥ 4 points, and 7% by ≥ 7 points.

CIBIC-plus scores significantly different between group 1 and placebo, $p < 0.05$.
A significantly higher proportion of participants in the low dose and high dose groups were considered improved (<4 points) (24% and 25% respectively) than the placebo participants (16%) (statistics not reported)

PDS one-quarter of high dose participants demonstrated clinically meaningful improvement (≥ 10%) compared to 15% in the placebo group ($p = 0.006$)

GDS scores significantly different between group 1 and placebo, $p < 0.05$

QoL (observed case analysis) change scores at 12 and 18 weeks (estimated from figure)	Rivastigmine 1–4 mg (n = 233)	Rivastigmine 6–12 mg (n = 231)	Placebo (n = 234)	p-Value drug versus placebo
ADAS-cog				
12 weeks	–1.8*	1.0†	–2.2	*$p < 0.05$, †$p < 0.001$
18 weeks	–2.2*	0.5†	–3.4	*$p < 0.05$, †$p < 0.001$
CIBIC-plus				
12 weeks	0.19	–0.08*	0.15	*$p < 0.05$
18 weeks	0.14	–0.13*	0.16	*$p < 0.05$
PDS				
12 weeks	–3.6	–0.77	–1.9	
18 weeks	–3.4	–0.76*	–4.0	*$p < 0.05$

Comments. CIBIC-plus figures appear to have labelled the Y axis minus numbers incorrectly.

continued

Adverse effects titration phase %	**Rivastigmine 1–4 mg**	**Rivastigmine 6–12 mg**	**Placebo**	***p*-Value versus placebo**
Sweating	2	6*	2	*$p < 0.05$
Fatigue	5	10*	4	*$p < 0.05$
Asthenia	2	10*	2	*$p < 0.05$
Weight decrease	1	4*	1	*$p < 0.05$
Malaise	1	3*	1	*$p < 0.05$
Allergy	2*	0	0	*$p < 0.05$
Hypertension	4*	3	1	*$p < 0.05$
Dizziness	15	24*	13	*$p < 0.05$
Somnolence	7	9*	2	*$p < 0.05$
Nausea	14	48*	11	*$p < 0.05$
Vomiting	7	27*	3	*$p < 0.05$
Anorexia	8*	20*	3	*$p < 0.05$
Flatulence	2	5*	1	*$p < 0.05$
Adverse events maintenance phase %	**Rivastigmine 1–4 mg**	**Rivastigmine 6–12 mg**	**Placebo**	***p*-Value versus placebo**
Dizziness	8	14*	4	*$p < 0.05$
Nausea	8*	20*	3	*$p < 0.05$
Vomiting	5*	16*	2	*$p < 0.05$
Dyspepsia	6*	5*	1	*$p < 0.05$
Sinusitis	1	4*	1	*$p < 0.05$
Mean change in body weight	no weight change	–3.91 Lbs (1.78 kg)	1.09 Lbs (0.50 kg)	

Comments. Mean body weight decreased for patients in group 2 but increased in the placebo group at all study evaluation times; the difference between these groups was significant (data not presented). Clinically notable weight changes (≥ 7%) were seen in more patients in the high dose group (21%, Fisher's exact test, $p < 0.001$) and the low dose (6%, $p < 0.05$) than in patients on placebo (2%).

Methodological comments

- Allocation to treatment groups: Randomised trial, all procedures managed by independent group. The research coordinator accessed an interactive voice response system that assigned the next available patient randomisation number, which maintained a blind assignment to medication.
- Blinding: Reports double-blinding, patients received 2 capsules twice daily with food – no details of placebo used. Unsure whether outcome assessors blinded.
- Comparability of treatment groups: More women than men enrolled; no demographic differences were found between treatment groups, with the exception of the overall population of enrolled men and women. Reports that no clinically meaningful differences in baseline medical conditions, or concomitant medications between treatment groups – data not reported.
- Method of data analysis: Efficacy analyses on several data sets, including intention to treat (all randomised patients); last-observation-carried-forward (randomised patients with ≥ one evaluation while on study medication); and observed cases (randomised patients with at ≥ least one evaluation while on study medication at designated assessment times). All comparisons to placebo were 2-tailed, with $p < 0.05$ statistically significant. Primary analyses for efficacy included ANOVA for CIBIC-plus and GDS; ANCOVA/ANOVA for ADAS-cog, PDS, and MMSE; and categorical analyses (Mantel–Haenszel) for ADAS-cog, CIBIC-plus, and PDS. Unsure whether the statistical significance level was corrected for multiple comparisons.
- Sample size/power calculation: The study population (~200 per group) was planned to achieve 90% power with alpha = 0.05 for detecting at least 3.0 points improvement on the ADAS-cog and an increase in the responder rate from 15% to 30% on the CIBIC-plus.
- Attrition/dropout: group 1) 199 (85%) completed trial, group 2) 149 (65%), group 3 196 (84%) (paper reports 197 but miscalculation). In group 3 high proportion withdrew due to adverse events, which may unmask blinding.

General comments

- Generalisability: Patients between 45–89 years, whose impairment was mild to moderately severe by MMSE.
- Outcome measures: No data on reliability and validity of measures given.
- Inter-centre variability: Unclear how many people undertook the clinical interviews between the centres.
- Conflict of interests: Novartis pharmaceuticals supported the project.

Quality criteria for Corey-Bloom *et al*.[57]

1. Was the assignment to the treatment groups really random?	Adequate
2. Was the treatment allocation concealed?	Adequate
3. Were the groups similar at baseline in terms of prognostic factors?	Reported
4. Were the eligibility criteria specified?	Adequate
5. Were outcome assessors blinded to the treatment allocation?	Unknown
6. Was the care provider blinded?	Partial
7. Was the patient blinded?	Partial
8. Were the point estimates and measure of variability presented for the primary outcome measure?	Adequate
9. Did the analyses include an intention-to-treat analysis?	Adequate
10. Were withdrawals and dropouts completely described?	Adequate

Reference and design	Intervention	Participants	Outcome measures
Author: Forette *et al*.[60] Year: 1999 Country: France Study design: RCT (multicentre) Number of centres: 11 (Belgium, Canada, France, Norway, UK) Funding: Novartis Pharma Corp	Treatment arms: (1) Rivastigmine twice a day (BiD), mean dose 9.6mg/day (see below for further dose information) (2) Rivastigmine three times a day (TiD), mean dose 10.1 mg/day (see below for further dose information) (3) Placebo Dose of active drugs: 10-week dose titration period followed by 8-week maintenance period. Dose started at 2 mg/day up to 12 mg/day with dose increments of 1 mg at day 4, 0.5 mg every fourth day until day 28, and 1 mg thereafter. Patients were titrated to their individual highest well-tolerated dose (≥ 6, 9 or 12 mg/day). Those who didn't tolerate doses of 6 mg/day were discontinued from the study. Mean dose tolerated was not statistically significant between groups ($p = 0.064$) Other interventions used: those who developed nausea	Number of participants: 114 enrolled; group 1) 45, group 2) 45, group 3) 24 Sample attrition/dropout: 15 withdrew due to adverse events in the titration phase (9 in the BiD group, 5 TiD group, 1 placebo), 2 others withdrew during titration phase (no reasons given). During the maintenance phase 10 withdrew due to adverse events (5 in the BiD group, 5 TiD group), 2 others withdrew during maintenance phase (no reasons given) Sample crossovers: none reported Inclusion/exclusion criteria for study entry: mild to moderate dementia (DSM-III-R), probable Alzheimer's dementia (NINCDS-ADRDA) with Mini-Mental State Examination (MMSE) scores of 12–26. Those with significant medical, neurological or psychiatric disorders were excluded Characteristics of participants: baseline characteristics only presented for those completing the study ($n = 70$): group 1) $n = 23$, group 2) $n = 28$, group 3) $n = 19$	Primary outcomes: Cognitive Function: Alzheimer's Disease Assessment Scale cognitive subscale (ADAS-cog); Weschler logical memory test (immediate and delayed recall); digit span test (forward and backward); word fluency. Clinical improvement: Clinicians' Interview-Based Impression of Change-plus (CIBIC-plus). Nurses' Observation Scale for Geriatric Patients (NOSGER) was used to assess cognitive functions and behaviour related to activities of daily living Secondary outcomes: Adverse events. Also compliance (counting unused capsules), overall tolerability, and vital signs (not extracted as not per protocol) Methods of assessing outcomes: cognitive function was assessed by an independent clinical neuropsychologist. CIBIC-plus information by interview by physician not given access to the results of psychometric tests or medical records after the first assessment at baseline. The NOSGER was completed by the next of kin or carer having most frequent contact with the patient

continued

Reference and design	Intervention	Participants	Outcome measures
	and/or vomiting were given domperidone 10–20 mg TiD to be taken 30 minutes before meals and always before study medication. When domperidone ineffective, metoclopromide 10 mg TiD before meals was the second choice. Where possible antiemetic therapy was stopped after 3 days, but was reinstated if nausea recurred	All means ± SD: ADAS-cog: group 1) 24.0 ± 11.6, group 2) 23.2 ± 8.5, group 3) 21.7 ± 8.8 Duration dementia (unsure times): group 1) 3.9 ± 2.2, group 2) 3.7 ± 2.7, group 3) 3.3 ± 2.1 Age (years): group 1) 69.5 ± 9.9, group 2) 71.7 ± 6.8, group 3) 72.5 ± 4.8 Total MMSE score: group 1) 19.8 ± 4.2, group 2) 19.4 ± 3.4, group 3) 19.2 ± 3.8	Patients attended clinic every 2 weeks. All assessments were repeated at 4 and 10 weeks of the titration period and at week 18 Length of follow-up: 18 weeks

Results

Outcomes	Rivastigmine BiD	Rivastigmine TiD	Placebo	*p*-Value
Cognition	**Rivastigmine BiD (*n* = 23)**	**Rivastigmine TiD (*n* = 28)**	**Placebo (*n* = 19)**	***p*-Value compared to placebo**
% responders CIBIC-plus	57*	36	16	*$p = 0.027$
Mean change ADAS-cog (estimated from figure)	–2.6	0.4	2.0	no significant differences
NOSGER	***n* = 23**	***n* = 27**	***n* = 19**	
Self-care	–0.4 ± 2.0	–0.6 ± 2.4	–0.3 ± 2.5	
Disturbing behaviour	–0.3 ± 2.1	–0.7 ± 3.4	0.1 ± 3.1	
Instrumental ADLs	0.4 ± 3.1	–0.7 ± 4.0	0.8 ± 4.0	
Memory	–0.7 ± 2.9*	–1.0 ± 2.7**	1.3 ± 3.7	*$p = 0.037$,
Mood	0.7 ± 3.0	–0.4 ± 3.4	–0.6 ± 3.2	**$p = 0.014$
Social behaviour	0.0 ± 2.6	–1.1 ± 3.8	0.3 ± 3.3	
Wechsler logical memory test (immediate recall)	1.8 ± 2.3	0.1 ± 2.3		BiD versus TiD $p = 0.012$

Comments. Reports that no significant differences were found between any of the treatment groups on the other psychometric parameters assessed, i.e. Weschler logical memory test (delayed recall); digit span test (forward and backward); and word fluency test.
ADAS-cog: no details scoring reported (decrease in score is improvement)
CIBIC-plus: global rating on scale 1–7, with very marked improvement 1, no change 4, very marked deterioration 7. Those given scores of 1–3 were considered responders.
NOSGER: no details scoring reported.

Adverse effects (incidence %)	Rivastigmine BiD (*n* = 45)	Rivastigmine TiD (*n* = 45)	Placebo (*n* = 24)	*p*-Value
Nausea	58	58	8	not reported
Vomiting	38	31	4	
Dizziness	27	9	0	
Anorexia	18	16	0	
Headache	16	20	4	

Comments: reports in text that 13 serious adverse events occurred. Only 2 of these occurred during the titration phase (weight loss and hypersalivation, and nausea and vomiting), and 2 during maintenance phase (bradypsychia, and abdominal pain) were considered by the investigator to be possibly related to study medication.

continued

Methodological comments

- Allocation to treatment groups: States patients were randomly assigned in a ratio of 2:2:1 to the three treatment groups. No description of concealment of allocation.
- Blinding: No details of placebo, no details of clinician blinding. States that cognitive function assessed by independent neuropsychologist but not described as blind to patients' treatment.
- Comparability of treatment groups: Baseline characteristics of total group not presented. No report of statistical analysis of baseline characteristics of those presented (those completing study) is given.
- Method of data analysis: Conducted an intent-to-treat population on safety measures, but an observed case analysis on efficacy measures. Analyses used pair-wise comparisons between treatment groups (Wilcoxon test). Additional analyses using Kruskall-Wallis for the NOSGER.
- Sample size/power calculation: Not reported
- Attrition/dropout: Some numbers withdrawing and reasons for withdrawal given but no details of numbers lost to follow-up.

General comments

- Generalisability: Minimal inclusion criteria, patients with mild to moderate dementia, probably AD with MMSE score of 12–26, minimal baseline characteristics given.
- Outcome measures: Doesn't report reliability or validity of psychological variables but does provide references for them.
- Inter-centre variability: Not reported.
- Conflict of interests: Funded by Novartis (pharmaceutical company).

Quality criteria for Forette *et al.*[60]

1. Was the assignment to the treatment groups really random?	Unknown
2. Was the treatment allocation concealed?	Unknown
3. Were the groups similar at baseline in terms of prognostic factors?	Unknown
4. Were the eligibility criteria specified?	Adequate
5. Were outcome assessors blinded to the treatment allocation?	Partial
6. Was the care provider blinded?	Unknown
7. Was the patient blinded?	Unknown
8. Were the point estimates and measure of variability presented for the primary outcome measure?	Adequate
9. Did the analyses include an intention-to-treat analysis?	Inadequate
10. Were withdrawals and dropouts completely described?	Partial

Reference and design	Intervention	Participants	Outcome measures
Author: Rösler *et al*.[58] Year: 1999 (including erratum) Country: Germany Study design: RCT, multicentre Number of centres: 45 (Austria, France, Germany, Switzerland, North America). Funding: Novartis Pharma	Treatment arms: (1) Rivastigmine 1–4 mg/day (low dose): see below for details (2) Rivastigmine 6–12 mg/day (high dose): see below for details (3) Placebo Dosages were increased weekly in steps of up to 1.5 mg/day during weeks 1–12 (dose escalation phase) and had to be within the target range by week 7. Decreases in dose were not permitted during this phase; however, if an adverse event occurred a dose could be omitted, maintained without increase for 2 weeks, or antiemetic drugs given. During weeks 13–26 (maintenance phase) doses could be increased or decreased within the assigned range with the aim of administering the highest dose that was well tolerated. The mean dose of rivastigmine was 3.7 (SD 0.59) mg/day in the low dose group and 10.4 (SD 2.13) mg/day in the high dose group. 190/210 (90%) of patients in the low dose group who were taking rivastigmine until the end of the study reached the maximum prescribed dose. This was 107/166 (64%) in the high dose group Other interventions used: approximately 81% ($n = 590$) were taking concomitant drug treatment. Mean number drugs taken per patient was 4.0 The most common drugs (taken by >10%) in each group were anti-infectives, cardiovascular,	Number of participants: 831 recruited, 106 excluded, 725 randomised. group 1) $n = 243$, group 2) $n = 243$, group 3) $n = 239$ Sample attrition/dropout: group 1) treatment discontinued 34 (withdrawal consent $n = 5$, failure to return $n = 3$, treatment failure $n = 1$, adverse event $n = 18$, death $n = 0$, non-compliance $n = 2$, other $n = 5$) Group 2) treatment discontinued 79 (withdrawal consent $n = 11$, failure to return $n = 2$, treatment failure $n = 2$, adverse event $n = 55$, death $n = 1$, non-compliance $n = 3$, other $n = 5$) Group 3) treatment discontinued 31 (withdrawal consent $n = 6$, failure to return $n = 2$, treatment failure $n = 2$, adverse event $n = 16$, death $n = 0$, non-compliance $n = 1$, other $n = 4$) Sample crossovers: none reported Inclusion/exclusion criteria for study entry: participants had to be 50–85 years and not able to bear children (older or younger people could enter the study with approval of the medical expert), Alzheimer's type dementia criteria (DSM-IV), probable AD (NINCD-ADRDA), scores 10–26 on Mini-Mental State Examination (MMSE). Each had a responsible caregiver. Those with concomitant disease such as hypertension, non-insulin dependent diabetes and arthritis were included. Those with severe and unstable cardiac disease severe obstructive pulmonary disease, or other life-threatening diseases (such as rapidly progressing malignancies) were excluded. Patients taking drugs for coexistent diseases were included except for those taking anticholinergic drugs, health food supplements containing acetylcholine precursors, putative memory enhancers, insulin and psychotropic drugs (the use of small doses of short acting benzodiazepines, chloral hydrate or haloperidol was allowed) Characteristics of participants: reports that demographic variables and disease characteristics were comparable at baseline across groups but does not present data. Only data presented: mean (range) ADAS-cog: group 1) 23.87 (4–60.7); group 2) 23.57 (5.7–58.0); group 3) 23.29 (3.3–57.8) mean (range) PDS: group 1) 53.8 (9.9–94); group 2) 55.22 (9.5–94.6); group 3) 54.1 (7.1–93.5) Presents data for total group ($n = 725$) only: females/male: 428 (59%)/297 (41%) mean age: 72 (range 45–95) years white: 97% ($n = 703$)	Primary outcomes: Alzheimer's Disease Assessment Scale – cognitive subscale (ADAS-cog), assesses memory, language, orientation, praxis, sourced by the patient; Clinician Interview Based Impression of Change plus caregiver information (CIBIC-plus), assesses global behaviour, general psychopathology, cognition, and activities of daily living (ADLs) sourced by patient and caregiver; Progressive Deterioration Scale (PDS), assesses ADLs (dressing and eating independently, social interaction, participation in housework and hobbies, awareness of time, handling of financial matters), sourced by the caregiver Secondary outcomes: MMSE and Global deterioration scale (GDS) as staging measures, Adverse events (coded using Sandoz medical technologies thesaurus). Also safety measures (not data extracted as per protocol) Methods of assessing outcomes: no description for psychological variables outcomes assessed at baseline, weeks 12, 18 and 26 or early withdrawal Length of follow-up: 26 weeks

continued

Reference and design	Intervention	Participants	Outcome measures
	gastrointestinal, respiratory, musculoskeletal, blood and nervous system disorders	mean duration dementia: 39 months mild disease: 41% (n = 298) moderate: 57% (n = 411) severe: 2% (n = 16) mean score MMSE: 19.9 (range 10–29) approximately 80% (n = 579) reported prior or current medical conditions or both. Mean number conditions per patient was 2.5	

Results

Outcomes	Low dose (n = 243)	High dose (n = 243)	Placebo (239)	p-Value versus placebo (ns unless stated)
Cognition ITT analysis (mean + 95% CI) at week 26				
ADAS-cog (NB. paper had scores neg not pos)	1.37 (2.27, 0.53)	–0.26 (0.66, –1.06)*	1.34 (2.19, 0.41)	*p = 0.011
No (%) with ≥ 4 point improvement	36/242 (15)	57/242 (24)	39/238 (16)	
CIBIC-plus	4.24 (4.02, 4.38)	3.91 (3.71, 4.09)*	4.38 (4.22, 4.58)	*p < 0.001
No (%) with improvement (scores 1,2 or 3)	69/233 (30)*	80/219 (37)**	46/230 (20)	*p < 0.05 **p < 0.001
PDS	–3.37 (–4.99, –1.61)	0.05 (–1.57, 1.77)*	–2.18 (–3.91, –0.49)	*p =0.07
No (%) with ≥ 10% improvement	45/241 (19)	70/241 (29)*	45/237 (19)	*p < 0.01
GDS	–0.22 (–0.3, –0.1)	–0.06 (–0.2, 0.0)*	–0.26 (–0.4, –0.2)	*p < 0.05
MMSE	–0.62 (–1.05, –0.15)	0.21 (–0.24, 0.64)*	–0.47 (–0.96, –0.04)	*p < 0.05
QoL observed case analysis (see comments below)				
CIBIC-plus (mean + 95% CI)				
Week 12	4.01 (3.83, 4.17)	3.88 (3.72, 4.08)	3.96 (3.83, 4.17)	
Week 18	4.06 (3.92, 4.28)	3.85 (3.7, 4.1)	4.09 (3.92, 4.28)	
Week 26	4.20 (3.99, 4.41)	3.93* (3.67, 4.13)	4.34 (4.09, 4.51)	*p < 0.05

Comments: LOCF analysis and OCs analysis not data extracted except where additional analyses undertaken for different time points. For ADAS-cog and PDS, data presented in figures 2 and 3 but unable to distinguish between each group in the figures.
ADAS-cog: scale from 0–70 where 0 = no errors (rarely achieved, even in general population) and 70 = severe impairment
CIBIC-plus: scale from 1–7, where 1, 2, 3 = marked, moderate, or minimal improvement, 4 = no change, and 5, 6, 7 = minimal, moderate or marked deterioration.
PDS: 29-item scale, scores range from 0 to 100. No description of how to interpret

Others				
Proportion withdrawing for any reason	14% (34/243)	33% (79/243)	13% (31/239)	see below

Comments: reports that significantly different between high dose group and low dose and placebo groups, no p-value

Adverse effects				
Proportion withdrawing for adverse events	7% (18/242*)	23% (55/242*)	7% (16/239)	see below

Comments: *states that adverse events evaluated in 242 in both treatment groups.
Reports that significantly different between high dose group and low dose and placebo groups, no p-value

continued

No. (%) adverse events occurring at 5% more often with Rivastigmine than in placebo or occurring with an incidence significantly different from placebo	**Low dose (*n* = 242)**	**High dose (*n* = 242)**	**Placebo (*n* = 239)**	***p*-Value compared with placebo**
Nausea	41 (17)*	121 (50)*	23 (10)	*$p < 0.05$
Vomiting	19 (8)	82 (34)*	14 (6)	
Dizziness	25 (10)	48 (20)*	17 (7)	
Headache	16 (7)	45 (19)*	18 (8)	
Diarrhoea	23 (10)	40 (17)*	21 (9)	
Anorexia	8 (3)	34 (14)*	4 (2)	
Abdominal pain	11 (5)	29 (12)*	7 (3)	
Fatigue	5 (2)	23 (10)*	6 (3)	
Malaise	3 (1)	23 (10)*	5 (2)	

Comments: overall significantly more patients reported at least one treatment-related adverse event in the high dose group (91% (220/242) than in the lower dose group (71% (172/242) or placebo (72% (172/239)

Methodological comments

- Allocation to treatment groups: Reports randomly allocated according to computer generated randomisation code at Novartis Pharma.
- Blinding: states double-blind. Capsules for treatment and placebo were identical and the number taken was the same at each dose in all groups. Not described how clinicians were blinded. High incidence of adverse events in high dose group may unblind patients and carers.
- Comparability of treatment groups: Reports no significant differences, but minimal data reported.
- Method of data analysis: Outcomes assessed as intention to treat, last observation carried forward (randomised patients with at least one evaluation while being treated), and observed cases (randomised patients with an evaluation made while on study drug at designated assessment times). Comparisons with placebo were two-tailed with $p < 0.05$. ANOVA and two-tailed pairwise students *t*-tests using the pooled error term from the ANOVA for CIBIC scale; ANCOVA and ANOVA with two tailed pairwise students *t*-tests using the pooled error term from the ANCOVA and ANOVA for the ADAS-cog and PDS; Mantel–Haenszel with blocking for centre for the ADAS-cog, the CIBIC scale and the PDS (categorical analyses). Fisher's exact test for adverse events.
- Sample size/power calculation: Study sample population of about 200 in each group was planned to enable achievement of 90% power with $\alpha = 0.05$ for detecting at least a 3.0 point improvement on the ADAS-cog and an increase from 15–30% among patients scoring <4 on the CIBIC scale.
- Attrition/dropout: Details of numbers and reasons for withdrawals. One patient from each treatment group lost to follow up in the adverse events data. Completed trial: group 1) $n = 209$, group 2) $n = 164$, group 3) $n = 208$

General comments

- Generalisability: Population appear to be generalisable to Alzheimer's patients with MMSE score 10–26.
- Outcome measures: Unclear how reliable and valid psychological variables are although references given.
- Inter-centre variability: Not reported.
- Conflict of interests: Funded by pharmaceutical company. Four authors are employees of the pharmaceutical company.

Quality criteria for Rösler *et al.*[58]

1. Was the assignment to the treatment groups really random?	Adequate
2. Was the treatment allocation concealed?	Unclear
3. Were the groups similar at baseline in terms of prognostic factors?	Unknown
4. Were the eligibility criteria specified?	Adequate
5. Were outcome assessors blinded to the treatment allocation?	Unknown
6. Was the care provider blinded?	Partial
7. Was the patient blinded?	Adequate
8. Were the point estimates and measure of variability presented for the primary outcome measure?	Adequate
9. Did the analyses include an intention-to-treat analysis?	Adequate
10. Were withdrawals and dropouts completely described?	Adequate

[Commercial/academic confidential information removed]

Appendix 9

Data extraction: galantamine RCTs

[Commercial/academic confidential information removed]

Reference and design	Intervention	Participants	Outcome measures
Author: Raskind *et al*.[61] Year: 2000 Country: USA Study design: RCT, double-blind, multicentre Number of centres: 33 Funding: Janssen Research Foundation (pharmacy company)	Treatment arms: 4 week single-blind placebo run-in period, then patients assigned to: (1) galantamine 24 mg/day: 8 mg/day for first week, followed by 16 mg/day in the second and 24 mg/day in the third. Then continued with the 24 mg/day dosage for an additional 5 months (2) galantamine 32 mg/day: 8 mg/day for first week, followed by 16 mg/day in the second and 24 mg/day in the third. Then increased to 32 mg/day for an additional 5 months (3) placebo Eligible patients then entered a 6-month open label extension phase: 8 mg/day for 1 week 16 mg/day for 1 week 24 mg/day for 5.5 months Other intervention used: Other antidementia medication had to be discontinued before entry to the study. The use of drugs for concomitant conditions was permitted during the study – except sedative-hypnotics and sedating cough and cold remedies, which were discontinued, if possible,	Number of participants: 764 patients screened, 128 excluded before or during run in, leaving 636 to be randomised (1) 212 to 24 mg/day maintenance dose (2) 211 to 32 mg/day maintenance dose (3) 213 placebo Sample attrition/dropout: During double-blind phase: (1) 68/212 discontinued (49 adverse events; 11 consent withdrawn; 3 non-compliance; 2 lost to follow-up; 3 other) (2) 89/211 discontinued (67 adverse events; 13 consent withdrawn; 4 non-compliance; 1 lost to follow-up; 4 other) (3) 41/213 discontinued (16 adverse events; 19 consent withdrawn; 2 non-compliance; 1 lost to follow-up; 3 other) Sample crossovers: none Inclusion/exclusion criteria for study entry: Inclusion criteria were as follows: a history of cognitive decline that had been gradual in onset and progressive over a period of at least 6 months; a diagnosis of probable AD according to the criteria of the NINCDS-ADRDA; presence of mild to moderate dementia: an MMSE score of 11 to 24 and a score of ≥ 12 on the standard cognitive subscale of the ADAS-cog Patients with stable and well-controlled concomitant medical disorders such as hypertension, heart failure (class I or class II), non-insulin-dependent diabetes	Primary outcomes: ADAS-cog/11; CIBIC-plus. Secondary outcomes: ADAS-cog/13 Proportion of responders on ADAS-cog 11 ADL inventory assessed using DAD scale Safety evaluations Adverse events Methods of assessing outcomes: ADAS-cog/11 with a score range of 0 to 70 CIBIC-plus scored by a trained clinician based on separate interviews with the patient and the caregiver. Scores ranged from 1 (markedly improved compared with baseline) to 7 (markedly worse). ADAS-cog/13 (score 0–85) Proportions of responders defined as improvement in ADAS/11 of ≥ 4 points compared to baseline. DAD based on interviews with the caregiver and assessed basic ADL, instrumental ADL, leisure activities, initiation, planning and organisation, and effective performance with 46 questions, with a score range of 0 to 100. Safety evaluations throughout the study comprised physical examinations, electrocardiogram, vital sign measurements and standard laboratory tests (data not extracted as per protocol) Monitoring for adverse events was recorded weekly for the first month of both the double-blind

continued

Reference and design	Intervention	Participants	Outcome measures
	48 hours before cognitive evaluation. Any other drugs with anticholinergic or cholinomimetic effects were avoided where possible	mellitus, and hypothyroidism were included. Patients were excluded if they had evidence of any neurodegenerative disorders other than AD, cardiovascular disease thought likely to prevent completion of the study, clinically significant cerebrovascular disease, active major psychiatric disorders, hepatic, renal, pulmonary, metabolic or urinary outflow obstruction, an active peptic ulcer, or any history of epilepsy, drug abuse or alcohol abuse. Patients who had been treated for AD with a cholinesterase inhibitor in the preceding 3 months were also excluded. All had a responsible caregiver Characteristics of participants: See table below	and open-label phases of the study, and at monthly intervals thereafter Assessments at 3 weeks (ADAS-cog) and 3 and 6 months Length of follow-up: 6 months (plus 6 month extension; data not extracted)

Results

Characteristics of patients	Galantamine 24 mg/day (*n* = 212)	Galantamine 32 mg/day (*n* = 211)	Placebo (*n* = 213)
Men/Women	73/139	87/124	82/131
Age, years*	75.9 ± 0.5	75.0 ± 0.6	75.3 ± 0.6
Weight, kg*	67.6 ± 1.0	67.3 ± 1.0	67.1 ± 1.0
White race, *n* (%)	195 (92.0)	190 (90.0)	196 (92.0)
Other medical conditions, *n* (%)	200 (94.3)	194 (91.9)	203 (95.3)
≥ APOE-ε4 allele, *n* (%)			
Time since probable AD diagnosed, years*	120 (60.0)	116 (61.7)	113 (58.2)
Total MMSE score*	1.02 ± 0.10	1.45 ± 0.13	1.13 ± 0.11
ADAS-cog/11 score*	19.5 ± 0.3	19.1 ± 0.3	19.2 ± 0.3
Total DAD score*	24.8 ± 0.7	25.8 ± 0.8	25.7 ± 0.8
	71.1 ± 1.5	70.3 ± 1.6	70.4 ± 1.6

Note: * Values are means ± SEM. Time since diagnosis only significant between the groups ($p = 0.02$) – unlikely to be clinically meaningful.

Outcomes

Cognition	Galantamine 24 mg/day	Galantamine 32 mg/day	Placebo	*p*-Value
ADAS-cog/11 score, mean (SEM) change from baseline after 6 months ITT (LOCF)	+1.9 (0.36)* (*n* = 202)	–1.4 (0.44)* (*n* = 197)	+2.0 (0.45) (*n* = 207)	*$p < 0.001$
CIBIC-plus, *n* (%) ITT (LOCF)	*n* = 186	*n* = 171	*n* = 196	
1 = markedly improved	3 (1.6)*	2 (1.2)**	1 (0.5)	*$p < 0.01$ **$p < 0.05$
2 = moderately improved	6 (3.2)*	4 (2.3)**	7 (3.6)	*$p < 0.01$ **$p < 0.05$
3 = minimally improved	28 (15.1)*	21 (12.3)**	19 (9.7)	*$p < 0.01$ **$p < 0.05$

continued

4 = no change	99 (53.2)*	91 (53.2)**	84 (42.9)	*$p < 0.01$ **$p < 0.05$
5 = minimally worsened	36 (19.4)*	43 (25.1)**	60 (30.6)	*$p < 0.01$ **$p < 0.05$
6 = moderately worsened	10 (5.4)*	9 (5.3)**	24 (12.2)	*$p < 0.01$ **$p < 0.05$
7 = markedly worsened	4 (2.2)*	1 (0.6)**	1 (0.5)	*$p < 0.01$ **$p < 0.05$

ADAS-cog/11 score, mean (±SEM) change from baseline over 6 months (OC) Estimated from figure	**Galantamine 24 mg/ Galantamine 24 mg**	**Galantamine 32 mg/ Galantamine 24 mg**	**Placebo/ Galantamine 24 mg**
3 weeks	–3.2	–2.0	–0.5
3 months	–3.3	–3.0	0
6 months	–2.2	–1.5	+2.0

Comments: At 6 months, OC analysis demonstrated a significant difference in the change in ADAS-cog/11 scores between galantamine- and placebo-treated patients. The differences in favour of galantamine were 3.9 points for the 24 mg/day and 3.8 points for the 32 mg/day groups ($p < 0.001$ in both cases). These differences were confirmed using the more conservative ITT analyses. The differences in change in ADAS-cog/11 scores between galantamine and placebo groups increased over time for both doses ($p < 0.001$).
ADAS-cog/11 responders: there were approximately twice as many ADAS-cog/11 responders in the galantamine-treated groups (33.3% galantamine 24 mg/day; 33.6% galantamine 32 mg/day) as in placebo (16.6%, $p < 0.01$ for both comparisons).
ADAS-cog/13: galantamine 24 mg/day and 32 mg/day produced a better outcome compared to placebo at 6 months; the treatment effect was 4.5 points at 24 mg/day and 4.1 points at 32 mg/day ($p <$ for both comparisons). This was confirmed on ITT analyses ($p < 0.01$ for all comparisons on both efficacy measures).
ADL inventory assessed using DAD scale: after 6 months of treatment, there were no significant differences between treatment groups in the mean change in total DAD score from baseline. Data not reported.

Adverse effects occurring at least 5% more frequently during either galantamine dose than with placebo during double-blind phase, *n* (%)	**Galantamine 24 mg/day (*n* = 212)**	**Galantamine 32 mg/day (*n* = 211)**	**Placebo (*n* = 213)**
Nausea	79 (37.3)	92 (43.6)	28 (13.1)
Vomiting	44 (20.8)	54 (25.6)	16 (7.5)
Dizziness	29 (13.7)	39 (18.5)	24 (11.3)
Diarrhoea	26 (12.3)	41 (19.4)	21 (9.9)
Anorexia	29 (13.7)	43 (20.4)	12 (5.6)
Weight loss	26 (12.3)	23 (10.9)	10 (4.7)
Abdominal pain	14 (6.6)	23 (10.9)	9 (4.2)
Tremor	11 (5.2)	7 (3.3)	1 (0.5)
Any adverse event	195 (92.0)	195 (92.4)	168 (78.9)

Comments: Withdrawals due to adverse events overall were 132/636 patients (21%); galantamine 24 mg/day 49/212 patients (23%); galantamine 32 mg/day 67/211 patients (32%); placebo 16/213 patients (8%). The majority of adverse events were mild to moderate in severity and predominantly gastrointestinal. The proportions of serious adverse events were comparable across treatment groups (13% to 16%). These included death in each group, neither related to treatment.

continued

Methodological comments

- Allocation to treatment groups: Random, states computer-generated code.
- Blinding: double-blind using identical single tablets taken twice daily. Investigators remained blind to the treatment to which patients were assigned.
- Comparability of treatment groups: Baseline demographics and medical characteristics were reported to be comparable. The only significant difference was time since diagnosis.
- Method of data analysis: All randomly assigned patients who took at least one dose of trial medication were included in the analyses of baseline characteristics and safety data. The primary analysis of 6-month efficacy data were based on patients who also provided baseline data for any of the ADAS-cog/11, CIBIC-plus, or DAD variables at designated assessment times – a traditional observed case (OC) analysis. To confirm robustness of the efficacy results, a more conservative 6-month ITT analysis was performed using the last-observation-carried-forward (LOCF) method. (For the extension, OC and ITT analyses were performed). All the results discussed in the study are based on OC analysis unless otherwise stated. ANOVA used for continuous variables, and the Cochran–Mantel–Haenszel test for categorical variables. ANCOVA model used in analysis of change from baseline score, with baseline ADAS-cog value as a covariate. The time-response relationship for change in ADAS-cog/11 was analysed using generalised linear interactive modelling. Exploratory ANOVA used to investigate any relationship between baseline characteristics.
- Sample size/power calculation: Uses data from previous study, which indicate that about 125 patients were needed in each treatment group to achieve 80% power ($\alpha = 0.025$ with a Bonferroni adjustment) to detect a difference of 2.75 points in the change in ADAS-cog/11 score between placebo and galantamine.
- Attrition/dropout: During double-blind phase: (1) 68/212 discontinued; (2) 89/211 discontinued; (3) 41/213 discontinued. Time of discontinuation not reported. Assume some dropped out early as not in the ITT population which included those with last observation carried forward.

General comments

- Generalisability: The study included patients with mild to moderate dementia (11–24 on MMSE) and a score of ≥ 12 on the ADAS-cog. Excluded patients on a number of grounds which may reduce generalisability to the normal AD population.
- Outcome measures: Outcome measures were appropriate and measured appropriately.
- Inter-centre variability: Not reported.
- Conflict of interests: Funding provided by Janssen Research Foundation.

Quality criteria for Raskind et *al.*[61]

1. Was the assignment to the treatment groups really random?	Adequate
2. Was the treatment allocation concealed?	Unclear
3. Were the groups similar at baseline in terms of prognostic factors?	Reported
4. Were the eligibility criteria specified?	Adequate
5. Were outcome assessors blinded to the treatment allocation?	Adequate
6. Was the care provider blinded?	Adequate
7. Was the patient blinded?	Adequate
8. Were the point estimates and measure of variability presented for the primary outcome measure?	Adequate
9. Did the analyses include an intention-to-treat analysis?	Inadequate
10. Were withdrawals and dropouts completely described?	Adequate

Reference and design	Intervention	Participants	Outcome measures
Author: Rockwood *et al*.[62] Year: 2001 Country: US, Canada, UK, SA, Australia, NZ Study design: RCT, multicentre Number of centres: 43 Funding: Janssen Research Foundation	Treatment arms: 4 week, single-blind, placebo run in phase, then 2:1 random assignment: (1) galantamine (gal) (2) placebo (pl) Gal patients received 8 mg/day for 1 week, increasing to 16 mg/day for the 2nd week and to 24 mg/day (12 mg twice daily) for the 3rd week. During week 4, dose could be increased to 32 mg (2×16) at the discretion of the investigator, based on tolerance. By the end of the 4th week, dose could be reduced from 32 to 24 mg/day, and patients continued with their final dose for a further 2 months. During week 4, 64 patients remained on 24 mg/day dose of galantamine, whereas 165 were increased to 32 mg/day dose, of whom 40 (24%) reverted to the lower dose during the week. Of the 125 patients remaining on the higher dose by the end of week 4, 103 (82%) completed the study. Of the patients who continued on or reverted to 24 mg/day dose during week 4, 72 (69%) completed the study Other interventions used: Protocol deviations occurred in 38 (10%) of randomised patients. 20 of these cases involved use of prohibited medications	Number of participants: 534 patients screened, 148 excluded before or during run-in, leaving 386 to be randomised. 125 to placebo, 261 to galantamine, of which 72 had final doses of 24 mg/day and 103 had final doses of 32 mg/day Sample attrition/dropout: 86/261 gal patients discontinued: 66 adverse events, 8 consent withdrawn, 3 non-compliance, 2 ineligible to continue, 2 lost to follow-up, 5 other reasons. 12/125 pl patients discontinued: 5 adverse events, 3 consent withdrawn, 2 ineligible to continue, 2 lost to follow-up Sample crossovers: None Inclusion/exclusion criteria for study entry: Probable AD according to NINCDS-ADRDA criteria. 1. Presence of mild to moderate dementia (11–24 on MMSE) AND ≥ 2 on ADAS-cog. Patients had to have regular contact with a responsible caregiver 2. Patients with concomitant diseases such as hypertension, congestive heart failure, non-insulin dependent diabetes mellitus and hypothyroidism were included in the study provided that the disease was controlled. 3. Patients with the following were excluded: other neurodegenerative disorders; cardiovascular disease likely to prevent completion of the study; clinically significant cerebrovascular, hepatic, renal, pulmonary, metabolic or endocrine conditions; clinically significant psychiatric disease, including moderate or severe or uncontrolled behavioural disturbances; urinary outflow obstruction; an active peptic ulcer; any history of epilepsy or significant drug or alcohol abuse. Patients previously treated with any cholinomimetic agent for AD, except muscarinic agonists, were excluded. 4. Any other medication being taken to treat AD had to be discontinued Characteristics of participants: Baseline characteristics Characteristic / gal (n = 261) / pl (n = 125) Males/females / 113/148 / 58/67 Age (y)* / 75.2 (0.45) / 74.6 (0.68) Weight (kg)* / 66.1 (0.86) / 68.5 (1.37) Smokers / 21 (8.0) / 9 (7.2)	Primary outcomes ADAS-cog11 to assess cognitive function, CIBIC-plus (clinician's impression plus caregiver input regarding clinical response): Secondary outcomes: expanded ADAS-cog13; proportions of responders (defined as improvements in ADAS-cog/11 ≥ 4 points from baseline); neuropsychiatric inventory (NPI), which assesses 10 domains of behavioural symptoms; disability assessment for dementia (DAD), based on an interview with the caregiver, to assess basic ADL, initiation, planning and organisation, performance and leisure. Also safety assessments (physical exams, ECGs etc) and recording of adverse events during first 2 weeks by investigator phoning patients/caregivers at weekly intervals. Safety was further evaluated at monthly clinic visits at weeks 3 and 4, and after 2 and 3 months. Sleep patterns assessed using Pittsburgh Sleep Quality Index (PSQI) Methods of assessing outcomes: CIBIC-plus was scored by a trained clinician based on separate interviews with the patient and the caregiver; clinician blinded to other assessments. Scores ranged from 1 to 7 (1 = marked improvement with respect to baseline, 7 = markedly worse) Length of follow-up: Efficacy assessments performed at baseline and after 1 and 3 months

continued

Reference and design	Intervention	Participants		Outcome measures
	The use of other concomitant medication was permitted, except that psychotropic drugs were discontinued 48 hours before cognitive evaluation where possible	ApoE ε4 genotype		
		Homozygous	38 (16.7)	14 (13.0)
		Heterozygous	111 (48.9)	56 (51.8)
		Total MMSE*	19.7 (0.24)	19.6 (0.32)
		ADAS-cog/11*	25.6 (0.65)	24.7 (0.85)
		Total NPI*	9.2 (0.66)	9.4 (1.01)
		Total DAD*	69.1 (1.42)	73.0 (1.91)
		Time since cognitive problem diagnosed (y)*	3.8 (0.20)	3.22 (0.19)
		Time since probable AD diagnosed (y)*	0.71 (0.07)	0.69 (0.1)
		Data are number (%) of pts, except those marked * which denotes mean (SE).		
		Proportion of pts taking concomitant medication was similar (89% pl, 88% gal). 33% (85/261) gal pts took concomitant psychotropic medications compared with pl group (25%, 31/125)		

AD, Alzheimer's disease; pt, patient; pts, patients; pl, placebo group; gal, galantamine group; g24, completed trial on 24 mg/day galantamine; g32, completed trial on 32 mg/day galantamine.

Results

Outcomes	Treatment X (*n* =)	Comparator X (*n* =)	*p*-Value
Cognition (Classic ITT)	**Galantamine 24–32 mg/day**	**Placebo**	***p*-Value**
ADAS-cog/11 (mean (SE) change from baseline)	–0.9 (0.31)** (*n* = 260)	+0.7 (0.47) (*n* = 125)	**$p < 0.01$
ADAS-cog/13 (mean (SE) change from baseline)	–1.1 (0.36)** (*n* = 258)	+0.7 (0.52) (*n* = 123)	**$p < 0.01$
No (%) ADAS-cog/11 responders ≥ 4 points improvement	73 (28.3) (*n* = 258)	27 (22.0) (*n* = 123)	
CIBIC-plus 9 No (%) patients in each category	(*n* = 248)	(*n* = 124)	
1 = markedly improved	1 (0.4)	0 (0)	
2 = moderately improved	7 (2.8)**	1 (0.8)	**$p < 0.01$
3 = minimally improved	56 (22.6)**	23 (18.5)	**$p < 0.01$
4 = no change	132 (53.2)**	54 (43.5)	**$p < 0.01$
5 = minimally worsened	43 (17.3)**	36 (29.0)	**$p < 0.01$
6 = moderately worsened	8 (3.2)**	9 (7.3)	**$p < 0.01$
7 = markedly worsened	1 (0.4)	1 (0.8)	
NPI (mean (SE) change from baseline	–0.4 (0.65) (*n* = 261)	+0.5 (0.64) (*n* = 125)	ns
DAD (mean (SE) change from baseline	–1.2 (0.83)** (*n* = 261)	–5.3 (1.17) (*n* = 125)	**$p < 0.01$
Comments: Expanded ADAS-cog13 – score range 0–85 Neuropsychiatric inventory (NPI) – score range 0–120 Disability assessment for dementia (DAD) – scale 0–100			
QoL (ITT LOCF)	**Galantamine 24–32 mg/day**	**Placebo**	***p*-Value**
ADAS-cog/11 (mean (SE) change from baseline)	–1.1 (0.33)** (*n* = 239)	+0.6 (0.45) (*n* = 120)	**$p < 0.01$
ADAS-cog/13 (mean (SE) change from baseline)	–1.2 (0.38)** (*n* = 239)	+0.7 (0.51) (*n* = 120)	**$p < 0.01$
No (%) ADAS-cog/11 responders ≥ 4 points improvement	72 (30.1) (*n* = 239)	27 (22.5) (*n* = 120)	

continued

CIBIC-plus 9 No. (%) patients in each category	(n = 240)	(n = 123)	
1 = markedly improved	1 (0.4)	0 (0)	
2 = moderately improved	7 (2.9)**	1 (0.8)	**$p < 0.01$
3 = minimally improved	53 (22.1)**	23 (18.7)	**$p < 0.01$
4 = no change	133 (55.4)**	53 (43.1)	**$p < 0.01$
5 = minimally worsened	40 (16.7)**	36 (29.3)	**$p < 0.01$
6 = moderately worsened	6 (2.5)**	9 (7.3)	**$p < 0.01$
7 = markedly worsened	0 (0)	1 (0.8)	
NPI (mean (SE) change from baseline)	–0.3 (0.7) (n = 241)	+0.5 (0.65) (n = 123)	
DAD (mean (SE) change from baseline)	–0.4 (0.76)*** (n = 241)	–5.2 (1.18) (n = 123)	***$p < 0.001$

QoL (OC)	**Galantamine 24–32 mg/day**	**Placebo**	***p*-Value**
ADAS-cog/11 (mean (SE) change from baseline)	–1.4 (0.4)** (n = 170)	+0.5 (0.42) (n = 108)	**$p < 0.01$
ADAS-cog/13 (mean (SE) change from baseline)	–1.6 (0.46)*** (n = 170)	+0.5 (0.49) (n = 106)	***$p < 0.001$
No (%) ADAS-cog/11 responders ≥ 4 points improvement	56 (32.9)* (n = 170)	21 (19.4) (n = 100)	*$p < 0.05$
CIBIC-plus 9 No. (%) patients in each category	(n = 170)	(n = 111)	
1 = markedly improved	1 (0.6)	0 (0)	
2 = moderately improved	7 (4.1)**	1 (0.9)	**$p < 0.01$
3 = minimally improved	41 (24.1)**	21 (18.9)	**$p < 0.01$
4 = no change	86 (50.6)**	48 (43.2)	**$p < 0.01$
5 = minimally worsened	30 (17.6)**	31 (27.9)	**$p < 0.01$
6 = moderately worsened	5 (2.9)**	9 (8.1)	**$p < 0.01$
7 = markedly worsened	0 (0.0)	1 (0.9)	
NPI (mean (SE) change from baseline	–0.7 (0.77) (n = 172)	0.0 (6.5) (n = 110)	
DAD (mean (SE) change from baseline	0.1 (0.87)** (n = 172)	–4.2 (1.16) (n = 110)	**$p < 0.01$

Comments

At 3 months, there was no difference between g24 and g32 patients in improvement from baseline ADAS-cog/11 during the fixed dose period (mean (SE) 1.4 (0.57), n = 99, and 1.5 (0.54), n = 71, ADAS points respectively). Findings supported by both ITT analyses. Gal better than pl on ADAS-cog/13 subscale (p = 0.004) and ADAS-cog/11 (p = 0.02). Overall clinical response measured by CIBIC-plus significantly better for gal than pl (p = 0.003), again backed-up by ITT analyses. Only 21% of the 170 gal patients deteriorated, compared with 37% of the 111 pl patients. At 3 months there was no significant change from baseline NPI score for either pl or gal patients

QoL (OC) At 1/3 months *(Numbers are estimated from Figs 2 and 3)*	**Galantamine 24–32 mg/day Months 1/3**	**Placebo Months 1/3**	***p*-Value**
ADAS-cog/11 (mean change from baseline)	–1.2*/–1.4**	–0.1/0.5	*$p < 0.05$ **$p < 0.01$
DAD (mean change from baseline)	0.1/0.1**	–1.4/–4.3	**$p < 0.01$

Comments:

Gal patients showed significantly better cognitive function than pl patients at 1 month (mean diff 1.1 points, $p < 0.05$) and at 3 months (1.9 points, p = 0.002) on the ADAS-cog/11 scale. Treatment differences were due to scores significantly improving from baseline in the gal patients ($p < 0.001$) while not changing significantly in the pl group.

Difference in mean change from baseline between gal and pl patients on DAD scale was 4.3 points (p = 0.004, also significant in ITT analyses). DAD score declined significantly from baseline in pl patients ($p < 0.001$ for OC and ITT analyses) but function was preserved in gal patients, whether in 32 or 24 mg/day groups (mean (SE) changes of 0.6 (1.21), n = 99, and –0.5 (1.24), n = 73, respectively)

continued

Others

QoL (OC) *(Numbers are estimated from Figure 4)*	**Galantamine 24–32 mg/day**	**Placebo**	***p*-Value**
DAD cluster (mean change from baseline at 3 months			$^{*}p < 0.05$ $^{**}p < 0.01$
Initiation	–0.25*	–4.5	$^{\dagger}p = 0.6$
Planning/organisation	0.75*	–2.9	
Performance	0**	–4.5	
Basic ADL	1.9*	–1.75	
Instrumental ADL	–1.25*	–6.5	
Leisure	0.8†	–6.9	

Comments:
Significant differences between pl and gal were also seen for each DAD cluster: initiation, planning, performance as well as both basic and instrumental ADL ($p \leq 0.05$ for all OC and LOCF-ITT analyses) except for DAD-leisure ($p = 0.06$ in OC analysis).

Adverse effects **Number (%) patients with adverse events occurring at least 5% more with gal than pl**	**Galantamine 24–32 mg/day (*n* = 261)**	**Placebo (*n* = 125)**
Nausea	84 (32.2)	14 (11.2)
Dizziness	39 (14.9)	5 (4.0)
Vomiting	38 (14.6)	5 (4.0)
Anorexia	31 (11.9)	3 (2.4)
Somnolence	20 (7.7)	1 (0.8)
Abdominal pain	18 (6.9)	2 (1.6)
Agitation	16 (6.1)	1 (0.8)
Any adverse event	225 (86.2)	79 (63.2)
Incidence of adverse events by phase of study (estimated from Fig 5)		
Week 1	18	10.5
Week 2	25	12.5
Week 3	28	15
Week 4	37	25
Week 5–12	57	48

Comments:
The incidence of adverse events during the dose escalation phase of the study in the gal group was greater than in the pl group. This difference was reduced during the maintenance phase. Proportion of serious adverse events was comparable (pl 6%, gal 8%). There were 2 deaths during the study, both in the placebo group. Discontinuations due to adverse events were more common in gal groups than pl group. Events most commonly associated with discontinuation: nausea (13%, 33/261), vomiting (6%, 15/261), dizziness (5%, 13/261) and anorexia (4%, 11/261).
Total PSQI scores did not change significantly in either treatment group.

Methodological comments

- Allocation to treatment groups: After a 4-week, single-blind, placebo run-in phase, patients were randomised to receive gal or pl in a 2:1 ratio using a computer generated code. The assignments were kept in sealed, opaque envelopes until the point of allocation.
- Blinding: Described as a double-blind trial. Investigator could increase dose at one point, so presumably was not blinded at that stage? CIBIC interviewer blinded to other assessments.
- Comparability of treatment groups: Baseline characteristics are described as 'comparable' but no *p*-values are presented. Proportion of patients taking concomitant medication was similar (89% pl, 88% gal). 33% (85/261) gal patients took concomitant psychotropic medications compared with pl group (25%, 31/125).

continued

- Method of data analysis: Primary efficacy analysis based on observed case (OC) analysis at 3 months. 3-month ITT analyses also conducted, along with LOCF. **All results discussed in the paper are based on OC analysis, unless otherwise stated**. Baseline characteristics compared using two way ANOVA for continuous variables and generalised Cochran–Mantel–Haenszel tests for categorical data. Van Elteren tests were used to test CIBIC-plus differences. Means and standard errors reported. Two graphs of changes over time are box and whisker plots, showing the upper and lower quartiles.
- Sample size/power calculation: 94 pl patients and 188 gal patients were required to achieve 80% power ($\alpha = 0.05$) for detecting a 2.5 point difference in the change in ADAS-cog/11 score between pl patients and each of the two gal dose groups.
- Attrition/dropout: 12/125 (9.6%) pl patients discontinued, 5/12 due to adverse events. 86/261 (33%) gal patients discontinued, 66/86 due to adverse events.

General comments

- Generalisability: Patients with probable AD and mild to moderate dementia (11–24 on MMSE) and a score of ≥ 2 on ADAS-cog. Patients with concomitant diseases such as hypertension, congestive heart failure, non-insulin dependent diabetes mellitus and hypothyroidism were included in the study provided that the disease was controlled. Patients with other neurodegenerative disorders; cardiovascular disease likely, clinically significant cerebrovascular, hepatic, renal, pulmonary, metabolic or endocrine conditions; clinically significant psychiatric disease, or certain other conditions were excluded.
- Outcome measures: Outcome measures were appropriate to the study area, and were measured appropriately.
- Inter-centre variability: Not discussed.
- Conflict of interests: Janssen Research Foundation funded the research. Two authors employed by Janssen Research Foundation.

Quality criteria for Rockwood *et al.*[62]

1. Was the assignment to the treatment groups really random?	Adequate
2. Was the treatment allocation concealed?	Adequate
3. Were the groups similar at baseline in terms of prognostic factors?	Reported
4. Were the eligibility criteria specified?	Adequate
5. Were outcome assessors blinded to the treatment allocation?	Partial
6. Was the care provider blinded?	Partial
7. Was the patient blinded?	Adequate
8. Were the point estimates and measure of variability presented for the primary outcome measure?	Adequate
9. Did the analyses include an intention-to-treat analysis?	Inadequate
10. Were withdrawals and dropouts completely described?	Adequate

Reference and design	Intervention	Participants	Outcome measures
Author: Tariot *et al*.[63] Year: 2000 Country: USA Study design: multicentre, parallel-group, placebo-controlled, double-blind trial Number of centres: Five Funding: Janssen Research Foundation	Treatment arms: 4-week placebo single-blind run in followed by randomisation to one of four groups for 5 months (weeks 1–21) (i) Galantamine 8 mg/day for 5 months (ii) Galantamine 8 mg/day for 4 weeks followed by galantamine 16 mg/day for 17 weeks (iii) Galantamine 8 mg/day for 4 weeks then galantamine 16 mg/day for 4 weeks and then a maintenance dose of galantamine 24 mg/day from weeks 9 to 21 (iv) Placebo for 5 months Galantamine and placebo were administered as identical single tablets taken orally twice daily Other interventions used: Other anti-dementia medication had to be discontinued before entry to the study, if licensed, and at least 30 days before entry if unlicensed. The use of drugs for concomitant conditions was permitted, with the exception of sedative-hypnotics and sedating cough and cold remedies, which were discontinued, if possible, 48 hours before cognitive evaluation. Any other drugs with anticholinergic or cholinomimetic effects were avoided	Number of participants: 1178 patients were screened and 978 were randomised to treatment (i) Galantamine 8 mg/day $n = 140$ (ii) Galantamine 16 mg/day $n = 279$ (iii) Galantamine 24 mg/day $n = 273$ (vi) Placebo $n = 286$ Sample attrition/dropout: 199 patients dropped out. (i) G. 8 mg/day $n = 32$ (ii) G. 16 mg/day $n = 60$ (iii) G. 24 mg/day $n = 61$ (vi) Placebo $n = 46$ Sample crossovers: No patients crossed over in this study Inclusion/exclusion criteria for study entry: History of cognitive decline, gradual in onset and progressive over at least 6 months; diagnosis of probably AD (NINCDS-ADRDA); MMSE 10–22, ADAS-cog score of ≥ 18 (from standard 11-item cognitive subscale). Concomitant diseases e.g. hypertension, heart failure (NY Heart Assoc. class I to II), type II diabetes mellitus, or hypothyroidism were allowed providing the illness was controlled. At inclusion a CT or MRI scan not older than 12 months had to be available showing no signs of clinically significant multi-infarct dementia or active cerebrovascular disease Exclusion criteria: patients with evidence of other neurodegenerative disorders; cardiovascular disease thought likely to prevent completion of the study; clinically significant psychiatric, hepatic, renal, pulmonary, metabolic or endocrine conditions, or urinary outflow obstruction; and active peptic ulcer; any history of epilepsy or significant drug or alcohol abuse. Patients treated for AD with a cholinomimetic agent in the preceding 60 days Characteristics of participants: (i) Galantamine 8 mg/day = G8 (ii) Galantamine 16 mg/day = G16 (iii) Galantamine 24 mg/day = G24 (vi) Placebo $n = 286$ = P ± values are mean ± SEM Men/Women: G8 50/90; G16 105/174; G24 90/183; P 108/178. Age, year: G8 76.0 ± 0.6; G16 76.3 ± 0.5; G24 77.7 ± 0.4; P 77.1 ± 0.5. Weight, kg: G8 70 ± 1.4; G16 68 ± 0.9; G24 67 ± 0.8; P 68 ± 0.8. White race, *n* (%): G8 132 (94); G16 260 (93); G24 249 (91); P 267 (93).	Primary outcomes: ADAS-cog 11-item subscale, score range 0–70 CIBIC-plus providing a global impression of patient deterioration or improvement Secondary outcomes: Proportion of responders as defined by the FDA (improvement in ADAS-cog of ≥ 4 points relative to baseline) Proportion of patients improved by ≥ 7 points on the ADAS-cog AD Cooperative Study Activities of Daily Living inventory (ADCS/ADL) to assess daily activities in patients with AD Neuropsychiatric Inventory (NPI) to assess the frequency and severity of symptoms in 10 behavioural domains Safety evaluations throughout the study were comprised of physical examinations, electrocardiography, vital signs, standard laboratory tests (not data extracted), and monitoring for adverse events (classified by WHO preferred terms) Methods of assessing outcomes: Data from one site were excluded from the efficacy analyses (but not the safety analyses) before the database was analysed, because the investigator failed to adhere to the principles of Good Clinical Practice. Of the 40 patients screened at this site, 32 patients were randomly assigned to the galantamine groups and 6 to the placebo group. CIBIC-plus: scored by a trained clinician, based on separate interviews with the patient and caregiver. Protocol recommended that the interview order should be standardised.

continued

Reference and design	Intervention	Participants	Outcome measures
		Other active medical conditions, *n* (%): G8 137 (98); G16 274 (98); G24 264 (97); P 274 (96). ≤ 1 APOE ε4 allele, *n* (%): G8 80 (62); G16 142 (55.9); G24 160 (64.5); P 165 (64.7). Time since cognitive problem diagnosed, year: G8 4.14 ± 0.21; G16 4.22 ± 0.16; G24 3.92 ± 0.16; P 4.33 ± 0.15. Time since probable AD diagnosed, year: G8 1.26 ± 0.12; G16 1.42 ± 0.11; G24 1.32 ± 0.11; P 1.42 ± 0.10. Total MMSE score: G8 18.0 ± 0.3; G16 17.8 ± 0.2; G24 17.7 ± 0.2; P 17.7 ± 0.2. ADAS-cog score: G8 27.8 ± 0.9; G16 29.4 ± 0.7; G24 29.0 ± 0.7; P 29.4 ± 0.6. ADCS/ADL score: G8 54.2 ± 1.2; G16 51.6 ± 0.9; G24 51.9 ± 1.0; P 52.3 ± 0.9. NPI score: G8 12.9 ± 1.2; G16 12.4 ± 0.8; G24 11.9 ± 0.8; P 11.0 ± 0.7 97% of all patients had active comorbid illnesses (mainly cardiovascular, musculoskeletal and ocular conditions). Most patients received concomitant medication: 97% in placebo gp, 96–98% in treatment gps. Antidepressant use: 27% in placebo gp, 26–34% in treatment gps. Other psychotropic medications (anxiolytics, hypnotics, neuroleptics): 23% placebo gp, 24–27% treatment gps	Assessments performed at weeks 4 & 13 and at 5 months Length of follow-up: 5 months

Outcomes	Galantamine 8 mg/day	Galantamine 16 mg/day	Galantamine 24 mg/day	Placebo
ADAS-cog: mean (SEM) change from baseline	*n* = 101	*n* = 208	*n* = 211	*n* = 225
5 months OC analysis	+0.1 (0.58)*	−1.5 (0.40)[‡ §]	−1.8 (0.44)[‡ ¶]	+1.8 (0.43)
Numbers in *italics* estimated from graph mean (SEM)				
1 month	*−0.8*	*−1.1*	*−0.9*	*−0.2*
3 months	*−0.8 (0.5)*	*−1.8 (0.3)*	*−1.7 (0.3)*	*0.6 (0.4)*
5 months (compare with 5 months above)	*0.1 (0.6)*	*−1.5 (0.4)*	*−1.8 (0.4)*	*1.8 (0.5)*
	n = 126	*n* = 253	*n* = 253	*n* = 255
5 months ITT analysis	+0.4 (0.52)	−1.4 (0.35)[‡ ¶]	−1.4 (0.39)[‡ ¶]	+1.7 (0.39)

* $p < 0.05$; † $p < 0.01$; ‡ $p < 0.001$ versus placebo; § $p < 0.05$; ¶ $p < 0.01$ versus 8 mg/day galantamine group.

Comments: ADAS-cog score range 0–70. OC analysis – treatment effects in favour of galantamine of 1.7 points (8 mg/day vs placebo $p < 0.05$), 3.3 points (16 mg/day $p < 0.001$) and 3.6 points (24 mg/day $p < 0.001$). Confirmed on ITT analysis for 16 mg/day and 24 mg/day groups ($p < 0.001$ for both comparisons), but not for the 8 mg/day group. Change in ADAS-cog from baseline was also significantly greater in the 2 higher dose groups than the lower dose group at 5 months ($p < 0.05$, 16 mg vs 8 mg; $p < 0.01$ 24 mg vs 8 mg). There was no significant difference between the 16 mg and 24 mg groups in mean change from baseline ADAS-cog at 5 months. All these results confirmed by the ITT analysis.
Values estimated from graph (in *italics*) – assume these are OC analysis although paper doesn't state explicitly. Not all SEMs estimable from graph as too overlapping.

continued

CIBIC-plus: Patients improved or no change, n (%)				
5 months OC analysis	54 (51)	143 (68)‡ §	136 (64)‡ ¶	112 (47)
5 months ITT analysis	68 (53)	169 (66)‡ §	162 (64)‡ ¶	128 (49)

Comments: Scores range from 1 = markedly improved compared with baseline, to 7 = markedly worse.
* $p < 0.05$; † $p < 0.01$; ‡ $p < 0.001$ versus placebo; § $p < 0.05$; ¶ $p < 0.01$ versus 8 mg/day galantamine group.

Proportion of responders at 5 months on ADAS-cog	Not reported	35.6%‡	37.0%‡	19.6%

Comments: Defined by the FDA as improvement in ADAS-cog of ≥ 4 points relative to baseline
‡ $p < 0.001$ versus placebo

Proportion patients, on ADAS-cog, improved (5 month OC analysis)	Not reported	15.9%†	22.3%†	7.6%

Comments: Defined by the FDA as improvement in ADAS-cog of ≥ 7 points relative to baseline. † $p < 0.01$ versus placebo. Confirmed by ITT analysis. The difference in the proportion of responders between the 24 mg/day and 16 mg/day groups approached significance on OC analysis ($p < 0.1$).

ADCS/ADL mean (SEM) change from baseline	*n* = 106	*n* = 212	*n* = 212	*n* = 235
5 month OC analysis	–3.1 (0.9)	–0.5 (0.6)‡ §	–1.6 (0.6)†	–4.0 (0.6)
Figures in *italics* estimated from graph mean (SEM)				
1 month	*–0.9*	*–0.1*	*–0.4*	*–0.9*
3 months	*–1.6 (0.7)*	*0.4 (0.5)*	*–0.7 (0.5)*	*–2.2 (0.5)*
5 months (compare with 5 months above)	*–3.2 (1.1)*	*–0.5 (0.5)*	*–1.6 (0.7)*	*–3.9 (0.7)*
	n = 129	*n* = 255	*n* = 253	*n* = 262
5 months ITT analysis	–3.2 (0.8)	–0.7 (0.5)‡ ¶	–1.5 (0.6)†	–3.8 (0.6)

Comments: Score range 0–78, was developed to assess daily activities, such as using household appliances, choosing clothes to wear, bathing and toileting. † $p < 0.01$; ‡ $p < 0.001$ versus placebo; § $p < 0.05$; ¶ $p < 0.01$ versus 8 mg/day galantamine group. There was no significant difference between the 16 mg/day and 24 mg/day groups on this measure. At 5 months patients' ADL were preserved in the G16 mg/day group, as indicated by a mean change in the ADCS/ADL score from baseline that was not significant (confirmed on OC and ITT analyses). The G16 mg/day regimen was associated with a significant smaller decrease in the ADCS/ADL score than the G8 mg/day regimen ($p < 0.05$ for OC analysis and $p < 0.01$, ITT analysis). Values estimated from graph (in *italics*) – assume these are OC analysis although paper doesn't state explicitly. Not all SEMs estimable from graph as too overlapping.

NPI mean (SEM) change from baseline	*n* = 106	*n* = 212	*n* = 212	*n* = 234
5 months OC analysis	2.3 (1.1)	–0.1 (0.8)*	–0.1 (0.9)*	2.3 (0.7)
Values in *italics* estimated from graph mean (SEM)				
1 month	*0.3 (0.7)*	*–0.3*	*–0.1*	*–1.1 (0.5)*
3 months	*2.0 (0.9)*	*–0.3 (0.5)*	*–0.1 (0.7)*	*0.9 (0.4)*
5 months (compare with 5 months above)	*2.3*	*–0.1 (0.9)*	*–0.1 (0.8)*	*2.3*
	n = 129	*n* = 255	*n* = 253	*n* = 262
5 months ITT analysis	2.3 (1.0)	–0.1 (0.7)*	0.0 (0.8)*	2.0 (0.7)

Comments: Score range 0 to 120, behavioural domains = delusions, hallucinations, agitation/aggression, dysphoria, anxiety, euphoria, apathy, disinhibition, irritability/lability and aberrant motor behaviour. * $p < 0.05$ from baseline. Values estimated from graph – assume these are OC analysis although paper doesn't state explicitly. Not all SEMs estimable from graph as too overlapping.

Withdrawals (number)	32	60	61	46
Adverse events	9	19	27	20
Non-compliance	4	7	10	3
Inefficacy	1	0	2	0
Ineligible	0	4	2	0
Others*	18	30	20	23

Comments: * The majority of discontinuations due to 'other' reasons were for withdrawal of consent.

continued

Adverse events number (%)				
Nausea	8 (5.7)	37 (13.3)	45 (16.5)	13 (4.5)
Vomiting	5 (3.6)	17 (6.1)	27 (9.9)	4 (1.4)
Anorexia	8 (5.7)	18 (6.5)	24 (8.8)	9 (3.1)
Agitation	21 (15.0)	28 (10.0)	22 (8.1)	27 (9.4)
Diarrhoea	7 (5.0)	34 (12.2)	15 (5.5)	17 (5.9)
Any adverse event	106 (75.7)	206 (73.8)	219 (80.2)	206 (72.0)
Any serious adverse event	14 (10.0)	28 (10.0)	35 (12.8)	31 (10.8)
Deaths	1 (0.7)	3 (1.1)	3 (1.1)	4 (1.4)

Comments: Adverse events listed above are those occurring at least 5% more often during treatment with any galantamine dose than with placebo. The majority of adverse events (inc. gastrointestinal symptoms) were mild in severity. There were few reports of muscle weakness in patients receiving galantamine (0.4 to 1.1%) and the incidence was similar in the placebo group (1.0%).

Methodological comments

- Allocation to treatment groups: Patients were randomised to one of four treatment arms using a computer-generated code. No further details given – not known if this code was dispensed over the phone, in envelopes, whether at a separate location or within one of the study sites. According to the randomisation ratio, half as many patients were assigned to the galantamine 8 mg group as the other three groups (see below).
- Blinding: Galantamine and placebo were administered as identical single tablets taken orally twice daily, which should have maintained blinding of both patient and caregiver. Assessment of CIBIC-plus was carried out by trained clinician but whether this person was also the treating physician (knowing details of adverse events etc) is unclear. No other details are given.
- Comparability of treatment groups: Reports that groups were comparable.
- Method of data analysis: All randomised patients who received at least one dose of trial medication were included in the analyses of baseline characteristics and safety data. The primary statistical analysis of efficacy was of observed cases (OC analysis). This included data from patients who were randomised and were available for evaluation at the designated assessment times. To confirm the robustness of the efficacy results, more conservative ITT analyses were performed using the last observation carried forward method (the last postbaseline observation available for each patient who received treatment). Note however that what the authors describe as ITT does NOT include all the randomised patients. Also, data from one site were excluded from the efficacy analyses (but not the safety analyses) before the database was analysed, because the investigator failed to adhere to the principles of Good Clinical Practice. Of the 40 patients screened at this site, 32 patients were randomly assigned to the 3 galantamine groups and 6 patients were randomly assigned to placebo. Comparisons of variable between each galantamine group and the placebo group were made with ANOVA for changes from baseline in ADAS-cog, ADCS/ADL, and NPI scores, including treatment and investigator as factors. An analysis of covariance (ANCOVA) model was also carried out in the analysis of change score, with baseline ADAS-cog value as covariate. ANCOVA and ANOVA models produced similar conclusions, therefore only the ANOVA results are reported. Treatment by investigator interaction was tested and removed from the model as it was not significant at the 5% level. Generalised Cochran–Mantel–Haensel tests were used to compare ADAS-cog response rates and Van Elteren tests for CIBIC-plus. For primary efficacy measures an *a priori* sequential step-down closed testing procedure was used to allow multiple statistical comparisons between each galantamine group and the placebo group, while maintaining the Type I error rate (α) at 0.05. Starting with the difference between the highest, 24 mg/day, dose and placebo, if the null hypothesis was rejected at the 0.05% levels (galantamine more effective than placebo) then the next highest dose against placebo was tested, and so on. If a lack of significance is found at the first or second step the testing procedure stops. The same method was used for exploratory comparisons between the two higher dose groups and the 8 mg/day group. Point estimates and SEMs were reported.
- Sample size/power calculation: Data from an earlier 6 month trial of galantamine indicated that 208 patients were needed in each treatment group to detect a mean difference of 3 points in the change from baseline in ADAS-cog score between patients in the placebo group and either of the 2 higher dose galantamine groups with >95% power ($\alpha = 0.05$). ADAS-cog scores were reported, changes in baseline scores between placebo and the 2 higher galantamine groups were over 3 points, results were significant. The galantamine 8 mg/day group was not powered to detect efficacy, but rather to contribute to the test for a dose response effect.
- Attrition/dropout: Information reported and seems comparable across groups although slightly fewer dropouts from placebo groups than others (16% vs 21–22%). 12–17% deviated from the protocol (similar across groups) but no details of what this means.

General comments

- Generalisability: Almost all patients (97%) had active comorbid illnesses (mainly cardiovascular, musculoskeletal, and ocular conditions). Most were also receiving concomitant medication, therefore the results of this study should be applicable to those with well-controlled additional illnesses. The study may not be applicable to patients with other neurological disorders, or more significant conditions as listed in exclusion criteria above.

continued

- Outcome measures: Outcome measures seem appropriate and relevant. Seem to have been measured appropriately (though unsure whether clinical assessor independent, blinding maintained as discussed above). Unsure as to the validity of ADCS/ADL.
- Inter-centre variability: Data from one site were excluded (38 patients) from the efficacy analyses (but not the safety analyses) before the database was analysed, because the investigator failed to adhere to the principles of Good Clinical Practice. The protocol recommended that the interview order should be standardised but does not report whether this was done/monitored. No other details regarding differences between centres or methods to standardise reporting are given.
- Conflict of interests: Janssen Research Foundation sponsored research and 3 authors are employees of Janssen Research Foundation

Quality criteria for Tariot *et al.*[63]

1. Was the assignment to the treatment groups really random?	Adequate
2. Was the treatment allocation concealed?	Unknown
3. Were the groups similar at baseline in terms of prognostic factors?	Reported
4. Were the eligibility criteria specified?	Adequate
5. Were outcome assessors blinded to the treatment allocation?	Unclear
6. Was the care provider blinded?	Adequate
7. Was the patient blinded?	Adequate
8. Were the point estimates and measure of variability presented for the primary outcome measure?	Adequate
9. Did the analyses include an intention-to-treat analysis?	Inadequate
10. Were withdrawals and dropouts completely described?	Adequate

Reference and design	Intervention	Participants	Outcome measures
Author: Wilcock *et al*.[64] Year: 2000 Country: UK Study design: Randomised, double-blind, parallel group, placebo controlled trial Number of centres: 86, in Canada, Finland, France, Germany, Norway, Sweden, the Netherlands and the UK Funding: Janssen Research Foundation (pharmaceutical company)	Treatment arms: After a 4-week, single-blind placebo run-in phase, patients were randomised to: (1) placebo (pl) group (2) gal24 group (3) gal32 group In both gal24 and gal32 groups, dose was 8 mg daily for 1st week, increasing to 16 mg/day for 2nd week, and to 24 mg/day for 3rd week. In 4th week, gal24 continued on 24 mg/day. And gal32 group were increased to 32 mg/day. Patients then continued with target dose for further 5 months Other interventions used: Overall, 85% (556/653) patients received concomitant drugs during the double-blind phase of the study, most commonly analgesics. 81 (38%) pl patients, 90 (41%) gal patients took concomitant psychotropic drugs during the double-blind phase	Number of participants: 753 enrolled, 100 excluded before or during run-in, and remaining 653 were randomised: gal24 $n = 220$, gal32 $n = 218$, pl $n = 215$ Sample attrition/dropout: Pl: 29/215 discontinued (19 adverse events, 4 non-compliance, 3 inefficiency, 3 other) and 186 (87%) completed the trial. Gal24: 44/220 discontinued (31 adverse events, 4 non-compliance, 1 inefficiency, 8 other) and 176 (80%) completed the trial. Gal32: 55/218 discontinued (48 adverse events, 1 non-compliance, 6 other) and 163 (75%) completed the trial. Patients who did and did not complete the study had comparable baseline characteristics, except that those that did not complete were older (mean age 74.1 v 71.7 years) Sample crossovers: None stated Inclusion/exclusion criteria for study entry: probable AD on NINCDS-ADRDA criteria and having mild to moderate dementia, defined as 11–24 on MMSE and a score of ≥ 12 on ADAS-cog/11 scale. Patients had to live with, or be visited at least 5 days a week by, a responsible caregiver. Patients with concomitant diseases such as hypertension, heart failure, type 2 diabetes mellitus and hypothydroidism were included in the study provided that their illness was well controlled. Patients were excluded if they had any other neurodegenerative disorder; multi-infarct dementia or clinically active cerebrovascular disease; cardiovascular disease thought likely to prevent completion of the study; clinically important cerebrovascular, psychiatric, hepatic, renal, pulmonary, metabolic or endocrine conditions or urinary outflow obstruction; an active peptic ulcer; or any history of epilepsy or serious drug or alcohol misuse. Patients who had been treated for AD with a cholinesterase inhibitor were also excluded. Any other drugs being taken to treat dementia had to be discontinued before the study. The use of drugs for other conditions was permitted during the study, except that sedative-hypnotic drugs and sedating cough and cold remedies were discontinued, if possible 48 hours before cognitive evaluation. Any other drugs with anticholinergic or cholinomimetic effects were avoided if possible Characteristics of participants: Baseline characteristics stated to be comparable, but no *p*-values provided in text.	Primary outcomes: ADAS-cog/11 to assess cognitive function; CIBIC-plus to provide global impression of patients' improvement or deterioration over time. Primary endpoint was at 6 months Secondary outcomes: Expanded ADAS-cog/13 (additional items are concentration/distractibility and delayed word recall); proportion of patients with improvements from baseline on ADAS-cog/11 of ≥ 0 and ≥ 4 points; Disability Assessment for Dementia (DAD) scale, based on an interview with the caregiver, to assess ADL. These assessments were performed at baseline and after 3 and 6 months. DAD scale was also measured after 3 weeks. Adverse events were recorded weekly during the first month by patient and/or caregiver, and at safety evaluations carried out at monthly clinic visits thereafter. Not data extracted as not per protocol. Subgroup analysis according to apolipoprotein E genotype – not data extracted as not per protocol Methods of assessing outcomes: Unclear whether patients were assessed by the same clinician (unlikely) Length of follow-up: 6 months

continued

Reference and design	Intervention	Participants	Outcome measures

Characteristic	Pl (*n* = 125)	Gal24 (*n* = 220)	Gal32 (*n* = 218)
Men/women	83/132	81/139	80/138
Mean (SD) age (year)	72.2 (7.6)	71.9 (8.3)	72.1 (8.6)
Mean (SD) weight (kg)	67.2 (12.1)	66.7 (12.8)	66.2 (13.4)
No (%) non-smokers	193 (90)	200 (91)	199 (91)
No (%) with apolipoprotein E4 allele*			
Homozygous	34 (18)	32 (17)	27 (15)
Heterozygous	83 (45)	97 (53)	95 (53)
Mean (SD) MMSE	19.3 (3.5)	19.5 (3.4)	19.0 (3.8)
Mean (SD) ADAS11	24.7 (9.3)	25.4 (9.4)	26.2 (10.4)
Mean (SD) DAD	66.6 (22.5)	69.9 (21.4)	69.6 (20.6)
Mean (SD) time since cognitive problem diagnosed (years)	3.5 (2.3)	3.6 (2.7)	3.7 (2.2)
Mean (SD) time since probable AD diagnosed (years)	0.8 (1.0)	0.9 (1.2)	0.8 (1.0)
Brain imaging findings (computed tomography or MRI findings in past 12 months)			
Territorial infarctions	1 (0.5)	7 (3)	2 (1)
Lacunar infarctions	17 (8)	10 (5)	16 (7)
White matter lesions	0	0	2 (1)
Tumour	0	1 (0.5)	1 (0.5)

* *n* = 185 pl, *n* = 184 gal24, *n* = 179 gal32

Abbreviations: AD, Alzheimer's disease; gal24, galantamine group with a max dose of 24 mg/day; gal32, galantamine group with a max dose of 32; pl, placebo group; OC, observed case analysis

Results

ADAS-cog/11 (negative change indicates improvement) (ITT)	**Gal 24**	**Gal 32**	**Placebo**	**Treatment difference from placebo (95% CI) Gal 24/Treatment difference from placebo (95% CI) Gal 32**
Number of patients	(*n* = 220)	(*n* = 217)	(*n* = 215)	
Mean (SE) change from baseline	–0.5 (0.38)	–0.8 (0.43)	2.4 (0.41)	2.9 (1.6 to 4.1) $p < 0.001$/ 3.1 (1.9 to 4.4) $p < 0.001$
No (%) with ≥ 0 points improvement	138 (63)	130 (60)	88 (41)	21.5 (12.0 to 31.0) $P < 0.001$/ 19.5 (10.0 to 29.0) $p < 0.001$
No (%) with ≥ 4 points improvement	64 (29)	70 (32)	32 (15)	14.0 (6.0 to 22.0) $p < 0.001$/ 17.0 (9.0 to 25.0) $p < 0.001$

continued

DAD score (negative change indicates deterioration)				
No. of patients	212	214	210	
Mean(se) change from baseline	–3.2 (1.02)	–2.5 (1.07)	–6.0 (1.08)	2.8 (–0.6 to 6.1) $p = 0.1$ (n/s)/ 3.4 (0.1 to 6.7) $p < 0.05$
CIBIC-plus				
Number of patients	206	198	203	
1 = much improved	0	0	0	
2 = moderately improved	7 (3)	9 (5)	1 (0.5)	
3 = minimally improved	29 (14)	39 (20)	32 (16)	<0.05/<0.001
4 = no change	91 (44)	82 (41)	68 (33)	
5 = minimally worsened	57 (28)	54 (27)	68 (33)	
6 = moderately worsened	17 (8)	14 (7)	32 (16)	
7 = much worsened	5 (2)	1 (1)	2 (1)	

Comments
ADAS-cog/11 – score range 0–70, with higher scores indicating greater cognitive impairment.
CIBIC-plus – 1 much improved, 4 = no change, 7 = much worse. Van Elteren test was used to test for differences in distribution of scores between pl and gal groups. NB. Unclear from paper whether *p*-value applies to all scores or just a score of 3?
ADAS-cog/13 – score range 0–85
DAD scale used 46 questions and had a score range of 0–100, with a higher score indicating better functioning.
Paper states that ITT but patient numbers not the same

Comments
Improvements in cognitive function from baseline in the gal groups were seen within one week of reaching a dose of 24 mg daily [mean 1.3 (SE 0.36) points for lower dose and 1.7 (0.37) for higher doses, both $p < 0.001$]. More gal patients (67–68%) improved or remained stable than pl patients (49%).
Extended ADAS-cog/13: treatment effect was 3.1 points for gal24 and 4.0 points for gal32 ($p < 0.001$ ITT and OC).
When both active treatment groups were combined for analysis, the difference between the pl and gal groups in the mean change from baseline disability assessment score was 3.18 points ($p < 0.05$).

Mean (SE) change from baseline in ADAS-cog/11 score over time, OC analysis	**Gal 24/**	**Gal 32**	**Placebo group**	***p*-Value**
1 month	–1.3/	–1.5*	–0.4	*$p < 0.05$
3 months	–2.1*** /	–2.4***	0.6	***$p < 0.001$
6 months	–0.8***	–1.6***	2.4	***$p < 0.001$

Comments
Numbers are estimated from Fig 2 (box and whisker plot)
Upper and lower quartiles shown on fig but hard to estimate

Mean (SE) change in ADAS-cog/11 score at 6 months according to baseline MMSE score, OC analysis	**Gal 24/**	**Gal 32**	**Placebo group**	***p*-Value**
<18	–0.7*** ($n = 45$)	–2.5*** ($n = 52$)	4.6 ($n = 49$)	***$p < 0.001$
≥18	–0.6** ($n = 111$)	–1.4*** ($n = 100$)	1.4 ($n = 122$)	**$p < 0.01$ ***$p < 0.001$

Comments
Numbers estimated from Fig 3. Both doses of galantamine were superior to pl on ADAS-cog/11 scale for patients with mild to moderate disease. Benefit was greatest for patients with moderately severe disease (baseline MMSE < 18), with a treatment difference between pl and gal32 or 7.0 points at 6 months ($p < 0.001$). NB. Described as 'exploratory analysis', presumably rather than planned subgroup analysis?

Adverse effects for which the difference between gal and pl groups was at least 5% – Number (%) patients	**Gal 24 ($n = 220$)**	**Gal 32 ($n = 218$)**	**Placebo ($n = 215$)**
Nausea	82 (37)	87 (40)	26 (12)
Vomiting	45 (20)	37 (17)	9 (4)
Diarrhoea	16 (7)	29 (13)	16 (7)
Dizziness	24 (11)	26 (12)	10 (5)

continued

Headache	21 (10)	25 (11)	7 (3)
Anorexia	22 (10)	23 (11)	0
Weight loss	17 (8)	11 (5)	1 (0.5)
Any adverse event	182 (83)	194 (89)	165 (77)

Comments:
At least 5% more galantamine patients than placebo patients reported nausea, vomiting, diarrhoea, dizziness, headache, anorexia and weight loss, with nausea being the most common adverse event. Nausea was rated as mild to moderate by most (153/169) patients. 115/169 galantamine patients who reported nausea had one episode, usually starting during dose escalation period. Median duration was 6 days for gal24 group and five days for gal32 group. 92% of adverse events were mild to moderate, and the proportion of serious adverse events was similar in the three treatment groups (12–13%). Events most commonly associated with discontinuation during galantamine treatment were nausea (10% (42/438) and vomiting (5% (24/438). 43/79 gal patients who discontinued due to adverse effects stopped during the dose escalation phase. Monthly discontinuation rates during the subsequent 5-month maintenance phase for galantamine groups (2.1% and 2.4%) were comparable to the discontinuation rate in the placebo group (2.1%).
Discontinuations due to adverse events were more common in galantamine (18%, 79/483) than placebo (9%, 19/215). More discontinued from high dose group than lower ((22%, 48/218) vs 14%, 31/220)).

Methodological comments

- Allocation to treatment groups: Randomisation schedule was computer generated at the Janssen Research Foundation. Assignments were kept in opaque, sealed, numbered envelopes, each containing the allocation for the next patient. Treatment was started on the day of allocation. The randomisation code was not broken until the database had been formally closed.
- Blinding: All doses were taken twice daily and were identical in appearance, taste and smell. For the CIBIC-plus, the clinician's interview was scored relative to baseline by a clinician blinded to other assessments and was based on separate interviews with the patient and the caregiver. Adverse events may 'unblind' patients, caregivers and clinicians.
- Comparability of treatment groups: Baseline characteristics were comparable.
- Method of data analysis: Primary analysis for efficacy data was based on traditional observed case (OC) analysis. If one item was missing from an assessment, that particular assessment was not included in the efficacy analysis. 6-month ITT analysis also performed, that included all randomised patients who had any efficacy assessment, whether at baseline or during treatment. Last observation was carried forward where actual data were not available. Changes from baseline were assessed using the 2-sided, paired *t*-test. Galantamine and placebo comparisons were made using: ANOVA, using treatment and country as factors, with pairwise Dunnett's tests for changes from baseline in ADAS-cog and DAD; generalised Cochran–Mantel–Haenszel test, controlling for country, for ADAS-cog/11; Van Elteren test (derived from CMH test), controlling for country, for CIBIC-plus; generalised linear mixed modelling used for testing time-response relation for change in ADAS-cog/11. All tests were evaluated at 5% significance. Serious protocol deviations were low (4%) so no per-protocol analysis was undertaken.
- Sample size/power calculation: Analysis of earlier trial data was used to calculate that 180 patients would be required in each treatment group to achieve 80% power ($\alpha = 0.025$) for detecting a 2.75 point difference in the change in ADAS-cog/11 scale after 6 months between pl patients and gal patients. This was the primary outcome measure.
- Attrition/dropout: Pl: 29/215 (13%) discontinued, gal24: 44/220 (20%) discontinued, gal32: 55/218 (25%) discontinued. Most discontinued due to adverse events.

General comments

- Generalisability: The study was based on patients with probable AD on NINCDS-ADRDA criteria who had mild to moderate dementia, defined as 11–24 on MMSE and a score of ≥ 12 on ADAS-cog/11 scale. Excluded patients with a number of concomitant diseases (see above).
- Outcome measures: Outcome measures were suitable for the type of study, and were assessed appropriately.
- Inter-centre variability: Tests for changes from baseline were controlled for country.
- Conflict of interests: The research was funded by Janssen Research Foundation. One author's department receives support from the pharmaceutical company and the author has received consultancy fees from pharmaceutical companies.

Quality criteria for Wilcock *et al.*[64]

1. Was the assignment to the treatment groups really random?	Adequate
2. Was the treatment allocation concealed?	Adequate
3. Were the groups similar at baseline in terms of prognostic factors?	Reported
4. Were the eligibility criteria specified?	Adequate
5. Were outcome assessors blinded to the treatment allocation?	Partial
6. Was the care provider blinded?	Adequate
7. Was the patient blinded?	Adequate
8. Were the point estimates and measure of variability presented for the primary outcome measure?	Adequate
9. Did the analyses include an intention-to-treat analysis?	Adequate (assume number discrepancy is an error)
10. Were withdrawals and dropouts completely described?	Adequate

Reference and design	Intervention	Participants	Outcome measures
Author: Wilkinson and Murray[65] Year: 2001 Country: UK Study design: RCT, multicentre Number of centres: 8 Funding: Shire Pharmaceuticals	Treatment arms: (1) Gal 18 mg/day (gal18) (2) Gal 24 mg/day (gal24) (3) Gal 36 mg/day (gal36) (4) Placebo (pl) 2 week washout period, then patients received 4 mg of galantamine twice daily; dose progressively increased at 2–3 day intervals until the target dosage levels of 6, 8 and 12 mg three times daily had been achieved (after 5, 8 and 14 days, respectively). Dose escalation period followed by 10 weeks of continuous fixed medication. Compliance was determined by returned tablet counts; patients using <60% of the prescribed medication were considered protocol violators Other interventions used: None stated	Number of participants: 285 randomised, gal18 $n = 88$, gal24 $n = 56$, gal36 $n = 54$, Pl $n = 87$ Sample attrition/dropout: gal18/gal24/gal36/pl Total discontinued 25 / 14 / 26 /14 Adverse events 19 / 10 / 24 / 8 Ineligible 0 / 0 / 0 / 2 Withdrew consent 4 / 2 / 1 / 2 Non-compliance 1 / 0 / 1 / 2 Other 1 / 2 / 0 / 0 Completed study: Pl 73/87 (84%) Gal18 63/88 (72%) Gal24 42/56 (75%) Gal36 28/54 (52%) Sample crossovers: none Inclusion/exclusion criteria for study entry: >45 years old; probable AD by NINCDS, ADRDA and DSM-III-R criteria of mild to moderate severity (MMSE 13–24). All participants were screened to exclude dementia secondary to causes other than AD, or any condition considered likely to interfere with the trial in the opinion of the investigator, and were required to have a resident relative or carer who could participate in the study and supervise medication. The use of the following medications was precluded: antidepressants, antipsychotic drugs, antiparkinsonian drugs, insulin, anticonvulsants, sedatives, antihypertensive agents (except ACE inhibitors and diuretics) and other centrally acting cholinergic or anticholinergic agents (except inhaled drugs for asthma) Characteristics of participants: Figs are mean ± SEM or *n* Characteristics gal18 gal24 gal36 pl Sex (M/F) % 44/56 41/59 43/57 41/59 Age (year) 72.7 ± 0.9 72.9 ± 1.1 75.4 ± 1.0 74.2 ± 0.9 MMSE score 18.8 ± 0.3 18.2 ± 0.4 18.8 ± 0.5 18.7 ± 0.3 ADAS-cog 26.0 ± 0.9 26.7 ± 1.1 25.7 ± 1.1 26.9 ± 1.0 AD duration (year) 3.1 ± 0.1 3.1 ± 0.2 3.9 ± 0.4 3.3 ± 0.2 *p*-Values not stated	Primary outcomes: ADAS-cog Secondary outcomes: CGIC and PDS-1 Routine clinical and physical examinations were carried out at each visit, and all adverse events were recorded. Serious adverse events were documented separately Pts assessed at baseline (week 0) and after 6 and 12 weeks of randomised treatment Methods of assessing outcomes: No background as to who undertook assessments Length of follow-up: 12 weeks

Abbreviations: pl, placebo group; gal18/24/32, galantamine groups with final doses of 18/24/32 mg/day respectively; pts, patients.

continued

Results

Cognition (ITT, LOCF)	**Gal18**	**Gal24**	**Gal36**	**Placebo**	***p*-Value**
ADAS-cog: *n*	*n* = 81	*n* = 55	*n* = 51	*n* = 82	
ADAS-cog: mean (SEM) change from baseline	–0.1 (0.7)	1.4** (0.9)	0.7† (0.7)	1.6 (0.7)	**$p < 0.01$ † $p = 0.08$
CGIC: *n*	*n* = 79	*n* = 53	*n* = 47	*n* = 83	
Much improved, *n* (%)	0 (0.0)	2 (3.8)	0 (0.0)	0 (0.0)	
Improved, *n* (%)	29 (36.7)	13 (24.5)	15 (31.9)	23 (31.3)	
No change, *n* (%)	38 (48.1)	29 (54.7)	26 (55.3)	34 (41.0)	
Worse, *n* (%)	12 (15.2)	9 (17.0)	6 (12.8)	23 (27.7)	
Much worse, *n* (%)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	
PDS-1: *n*	*n* = 88	*n* = 56	*n* = 54	*n* = 87	
Much worse, *n* (%)	4 (4.5)	1 (1.8)	1 (1.9)	7 (8.0)	
Worse, *n* (%)	12 (13.6)	5 (8.9)	11 (20.3)	15 (17.2)	
No change, *n* (%)	61 (69.3)	42 (75.0)	38 (70.4)	57 (65.5)	
Improved, *n* (%)	9 (10.2)	7 (12.5)	4 (7.4)	8 (9.2)	
Much improved, *n* (%)	2 (2.3)	1 (1.8)	0 (0.0)	0 (0.0)	

Comments:
ADAS-cog: no scoring details provided, but reference given.
CGIC scores, graded from 1 to 7, were collapsed into a 5-point scale: much improved, improved, no change, worse, and much worse.
PDS-1: activities were rated on a 10-cm visual analogue scale (VAS) (0 = worst score, 10 = best score). Total VAS score (sum of 27 questions) then evaluated.
At 12 weeks, gal24 group was significantly better than placebo (3 points on scale, $p = 0.01$) on ITT analysis of change in ADAS-cog score from baseline.

QoL (PP)	**Gal18**	**Gal24**	**Gal36**	**Placebo**	***p*-Value**
ADAS-cog: *n*	*n* = 62	*n* = 44	*n* = 29	*n* = 53	
ADAS-cog: mean (SEM) change from baseline	–0.8* (0.8)	1.9** (1.0)	1.8** (0.9)	2.3 (0.9)	*$p < 0.05$ **$p < 0.01$ † $p = 0.08$
CGIC: *n*	*n* = 61	*n* = 44	*n* = 29	*n* = 74	
Much improved, *n* (%)	0 (0.0)	2 (4.5)	0* (0.0)	0 (0.0)	*$p < 0.05$
Improved, *n* (%)	27 (44.3)	13 (29.5)	14* (48.3)	25 (33.8)	*$p < 0.05$
No change, *n* (%)	24 (39.3)	21 (47.7)	13* (44.8)	28 (37.8)	*$p < 0.05$
Worse, *n* (%)	10 (16.4)	8 (18.2)	2* (6.9)	21 (28.4)	*$p < 0.05$
Much worse, *n* (%)	0 (0.0)	0 (0.0)	0* (0.0)	0 (0.0)	*$p < 0.05$
PDS-1: *n*	*n* = 62	*n* = 44	*n* = 29	*n* = 74	
Much worse, *n* (%)	1 (1.6)	0* (0.0)	1 (3.4)	3 (4.1)	*$p < 0.05$
Worse, *n* (%)	9 (14.5)	2* (4.5)	5 (17.2)	13 (17.6)	*$p < 0.05$
No change, *n* (%)	43 (69.4)	34* (77.2)	19 (16.5)	50 (67.6)	*$p < 0.05$
Improved, *n* (%)	7 (11.3)	7* (15.9)	4 (13.8)	8 (10.8)	*$p < 0.05$
Much improved, *n* (%)	2 (3.2)	1* (2.3)	0 (0.0)	0 (0.0)	*$p < 0.05$

Comments:
At 12 weeks, all galantamine groups were significantly better than placebo using per protocol (PP) analysis of change in ADAS-cog score from baseline. The largest effect was for gal24 (4.2 points)
Although not powered to detect significant differences in secondary outcomes, PP analysis shows gal24 had a significantly better improvement in PDS-1 score than the placebo group. On the CGIC scale, there was a significant difference between gal36 and pl patients.

continued

Adverse effects reported by ≥5% of patients: *n* (%)	**Gal18 (*n* = 88)**	**Gal24 (*n* = 56)**	**Gal36 (*n* = 54)**	**Placebo (*n* = 87)**	***p*-Value** Not reported
Vomiting	15 (17.0)	4 (7.1)	9 (16.7)	4 (4.6)	
Nausea	15 (17.0)	10 (17.9)	20 (37.0)	3 (3.4)	
Headache	5 (5.7)	6 (10.7)	8 (14.8)	4 (4.6)	
Diarrhoea	2 (2.3)	3 (5.4)	2 (3.7)	2 (2.3)	
Decreased appetite	5 (5.7)	2 (3.6)	4 (7.4)	2 (2.3)	
Dizziness	4 (4.5)	2 (3.6)	4 (7.4)	3 (3.4)	
Any adverse event	49 (55.7)	33 (58.9)	38 (70.4)	38 (43.7)	
Withdrawn as a result of adverse events	19 (21.6)	10 (17.9)	24 (44.4)	8 (9.2)	
Number reporting serious adverse events	6 (6.8)	0 (0.0)	5 (9.3)	3 (3.4)	

Comments:
Dose-related side effects, principally cholinergic, were mainly mild and transient and occurred predominantly during the initial 5-day to 2-week dose escalation phase. Thereafter, the incidence of side effects declined rapidly to a level similar to placebo.

Methodological comments

- Allocation to treatment groups: A computer generated randomisation code was used to allocate patients to one of the four treatment groups, the patients remained blind to this and all interim analyses and the sealed codes were verified at the end of the study. *An earlier statistical examination of data from 163 patients who had completed the trial (second interim analysis) indicated that predetermined conditions for discontinuation (i.e. efficacy and/or tolerability criteria) had been satisfied in two of the three active treatment arms (24 and 36 mg/day). Subsequent recruitment was therefore continued only into the placebo and 18-mg/day groups. Investigators and all personnel directly involved in the study remained blinded to this fact.*
- Blinding: To maintain double-blind conditions, galantamine and placebo tablets were identical in appearance.
- Comparability of treatment groups: Baseline characteristics appear to be similar, although no *p*-values are presented.
- Method of data analysis: A group sequential design was used in an interim statistical analysis of the change in ADAS-cog. Statistical analysis was structured to detect a standardised difference of 0.5 (ratio difference/variation) with 90% power and a significance level of 0.05 (two-tailed), equating to a 2-point difference in change from baseline ADAS-cog score between placebo and galantamine groups. Interim analyses were carried out after approximately 20 patients had completed assessment in each group. ITT using LOCF included those patients who were randomised into the trial and who subsequently received at least one dose of trial medication. Per protocol (PP) analysis included patients in the ITT population who were not protocol violators (PP was only to be carried out if it included 60–95% of the ITT population). ANOVA used to test for ADAS-cog differences from baseline between the four treatment groups, with Dunnett's test for pairwise comparison among the four treatment groups. CGIC and PDS-1 differences between groups were analysed using the Cochran–Mantel–Haeszel test. 5% significance used in all cases. Results are shown as mean ± SEM or ± 95% confidence intervals.
- Sample size/power calculation: It was estimated that 240 evaluable patients were needed and that 240–360 patients would need to be enrolled.
- Attrition/dropout: Patients using <60% of the prescribed medication were considered protocol violators. Dropout rates: Pl (16%), Gal18 (28%), Gal24 (25%), Gal36 (48%). Withdrawals were most commonly due to adverse events. Drop out rate for gal36 group is very high.

General comments

- Generalisability: The study was limited to patients aged over 45 with standard criteria for MMSE scores. Patients were screened to exclude dementia secondary to causes other than AD, or 'any condition considered likely to interfere with the trial in the opinion of the investigator'. The particular concomitant conditions which formed inclusion/exclusion criteria are not listed.
- Outcome measures: Outcome measures were appropriate.
- Inter-centre variability: not discussed.
- Conflict of interests: Funded by Shire Pharmaceuticals.

Quality criteria for Wilkinson and Murray[65]

1. Was the assignment to the treatment groups really random?	Adequate
2. Was the treatment allocation concealed?	Adequate
3. Were the groups similar at baseline in terms of prognostic factors?	Reported
4. Were the eligibility criteria specified?	Adequate
5. Were outcome assessors blinded to the treatment allocation?	Unknown
6. Was the care provider blinded?	Adequate
7. Was the patient blinded?	Adequate
8. Were the point estimates and measure of variability presented for the primary outcome measure?	Adequate
9. Did the analyses include an intention-to-treat analysis?	Inadequate
10. Were withdrawals and dropouts completely described?	Adequate

Reference and design	Intervention	Participants	Outcome measures
Author: Wilkinson *et al*.[66] (conference proceeding) Year: 2000 Country: UK Study design: RCT Number of centres: 43 Funding: not reported	Treatment arms: Following a 4 week single-blind run in period, patients received: (1) galantamine, dose escalated to 24 mg/day over 3 weeks. During week 4 the galantamine could be increased to 32 mg/day at the discretion of the investigator, based on how well the patient tolerated the drug. By the end of the 4th week, the investigator could reduce the galantamine dose from 32 mg/day to 24 mg/day based on tolerability. Thereafter the dose was fixed (2) placebo, no details, although states that during week 4 the placebo could be increased. Other interventions used: not reported	Number of participants: 386 patients randomised, group 1) galantamine = 261, group 2) placebo = 125 Sample attrition/dropout: 25% did not complete study, no further details Sample crossovers: none Inclusion/exclusion criteria for study entry: probable AD (NINCDS-ADRDA), mild-to-moderate dementia (MMSE 11–24 and score ≥ 12 on the ADAS-cog), no clinical evidence of another cause for the cognitive impairment Characteristics of participants (number (%) unless stated: Males/females: group 1) 113/148; group 2) 58/67 Mean ± SEM Age (years): 1) 75.2 (0.45); 2) 74.6 (0.68) Smokers: 1) 21 (8.0); 2) 9 (7.2) Other active medical conditions: 1) 235 (90.0); 2) 112 (89.6) Total MMSE score (mean ± SEM): 1) 19.7 (0.24); 2) 19.6 (0.32) ADAS-cog (mean ± SEM): 1) 25.6 (0.65); 2) 24.7 (0.85) Total DAD score (mean ± SEM): 1) 69.1 (1.42); 2) 73.0 (1.91) Time since diagnosis (mean ± SEM years): 1) 0.71 (0.07); 2) 0.69 (0.1)	Primary outcomes: ADAS-cog, CIBIC-plus, Disability Assessment for Dementia (DAD) scale Secondary outcomes: adverse events. Safety (physical examinations, ECG, vital signs, laboratory tests), not data extracted as per protocol Methods of assessing outcomes: not reported Length of follow-up: 3 months

continued

Results (all LOCF unless stated)

Outcomes	Galantamine	Placebo	*p*-Value vs placebo
ADAS-cog change from baseline, mean ± SEM	–1.1 (0.33), $n = 239$	0.6 (0.45), $n = 120$	$p < 0.01$
CIBIC-plus (%)	$n = 240$	$n = 123$	
1 (markedly improved)	0.4	0	
2 (moderately improved)	2.9	0.8	
3 (minimally improved)	22.1	18.7	Overall $p < 0.01$
4 (no change)	55.4	43.1	
5 (minimally worsened)	16.7	29.3	
6 (moderately worsened)	2.5	7.3	
7 (markedly worsened)	0	0.8	
DAD (mean ± SEM) change from baseline	–0.4 (0.76)	–5.2 (1.18)	$p < 0.001$
DAD cluster scores (mean ± SEM), estimated from figure			
Initiation	–0.1	–4.5	$p < 0.05$
Planning/organisation	0.6	–3.0	$p < 0.05$
Performance	0.0	–4.5	$p < 0.01$
Basic	1.7	–1.9	$p < 0.05$
Instrumental	–1.5	–6.5	$p < 0.05$
Leisure	0.9	–7.0	$p = 0.06$

Comments: in the galantamine group the improvement from baseline on the ADAS-cog was seen regardless of whether patients were maintained on a dose of 32 mg/day (mean improvement of 1.4 points) or 24 mg/day (mean improvement 1.5 points), on the observed case analysis. Similarly in the DAD scale (figures not presented).

Adverse effects

Comments: most AEs in the galantamine group were gastrointestinal in origin, mild-to-moderate in severity and were mainly confined to the dose-escalation phase. The AEs most frequently (≥ 5%) associated with discontinuations during galantamine therapy were nausea (13%), vomiting (6%) and dizziness (5%). The proportion of serious AEs was comparable among treatment groups (6% placebo, 8% galantamine).

Methodological comments

- Allocation to treatment groups: States patients randomised to receive galantamine or placebo in a 2:1ratio. No method of randomisation presented.
- Blinding: Not reported as a double-blind study. no details of blinding of outcome assessors.
- Comparability of treatment groups: Both groups had similar baseline characteristics.
- Method of data analysis: The primary analysis was an observed cases (OC) analysis. A last observation carried forward (LOCF) analysis, using the last post-baseline observations available for each patient who received treatment, was also used. All results presented here are LOCF unless stated otherwise.
- Sample size/power calculation: Not reported.
- Attrition/dropout: Reports that 75% completed the study; no details given in individual groups.

General comments: Study presented as a conference poster, therefore limited detail.

- Generalisability: Those with mild-to-moderate AD.
- Outcome measures: Appropriate.
- Inter-centre variability: Not reported.
- Conflict of interests: Not reported.

Quality criteria for Wilkinson[66]

1. Was the assignment to the treatment groups really random?	Unclear
2. Was the treatment allocation concealed?	Unclear
3. Were the groups similar at baseline in terms of prognostic factors?	Reported
4. Were the eligibility criteria specified?	Adequate
5. Were outcome assessors blinded to the treatment allocation?	Unclear
6. Was the care provider blinded?	Inadequate
7. Was the patient blinded?	Inadequate
8. Were the point estimates and measure of variability presented for the primary outcome measure?	Adequate
9. Did the analyses include an intention-to-treat analysis?	Inadequate
10. Were withdrawals and dropouts completely described?	Inadequate

Appendix 10

Data extraction: head-to-head comparisons

Reference and design	Intervention	Participants	Outcome measures
Author: Fuschillo *et al.*[68] Year: 2001 Country: Italy Study design: RCT Number of centres: 1 Funding: not reported	Treatment arms: (1) donepezil 5 mg/day in the evening (2) rivastigmine 1.5 mg/day for one week in the evening, and then doses increased weekly by steps of 1.5 mg/day up to the dose range of 6–9 mg/day, in two daily administrations, if tolerated. Other interventions used: patients were allowed to continue most medications for co-existent diseases, except for anticholinergic drugs, Ach-precursors, and other psychotropic drugs	Number of participants: 27; donepezil $n = 16$, rivastigmine $n = 11$ Sample attrition/dropout: not reported Sample crossovers: none Inclusion/exclusion criteria for study entry: patients were those consecutively referred to a neuropsychogeriatric ward meeting criteria for AD (DSM-IV and NINCDS-ADRDA). MMSE between 10–21. CT or MRI consistent with AD in the 12 months preceding inclusion in the study. Patients with concomitant diseases were included unless the medical condition was severe and/or unstable. Each patient had a reliable and responsible caregiver Characteristics of participants (mean ± SD, or *n* (%)): Age, years (range): Donepezil 68.1 ± 5.6 (54–77); Rivastigmine 66.2 ± 9.2 (53–77) Gender Male/female: Don 7(44)/9(56); Riv 5(45)/6(55) Education (years) 0–5/6–8/9–14: Don 9(56)/5(32)/2(12); Riv 6(55)/3(27)/2(18) Dementia duration, months (range): Don 21.4 ± 7.8 (11–35); Riv 22.4 ± 9.1 (10–36) ADAS-cog: Don 43.0 ± 7.6; Riv 40.3 ± 6.7 MMSE: Don 13.7 ± 3.4; Riv 13.2 ± 3.3 PSMS: 9.2 ± 2.9; Riv 11.5 ± 3.9	Primary outcomes: MMSE, ADAS-cog, Physical Self Maintenance Scale (PSMS) of the ADL test Secondary outcomes: adverse events. Safety evaluations including physical examinations, ECG, vital signs, laboratory evaluations, (not data extracted as per protocol) Methods of assessing outcomes: ADAS-cog and MMSE information source was the patient. The PSMS information source was the caregiver (see below for scoring details) Adverse events were gathered from the patients' and/or caregivers' reports Assessments were carried out at baseline and at weeks 6, 12, 18, 24, and 30 Length of follow-up: 30 weeks

continued

Results

Outcomes	Donepezil (n = 16)	Rivastigmine (n = 11)	p-Value
ADAS-cog (mean ± SD)			Reports that overall ns but assume this is within group not between group comparison
Week 6	41.7 ± 7.3	39.2 ± 6.0	
Week 12	40.8 ± 6.7	38.8 ± 5.9	
Week 18	40.4 ± 6.8	38.0 ± 6.1	
Week 24	39.7 ± 6.7	36.8 ± 5.9	
Week 30	39.4 ± 6.6	36.5 ± 5.7	
MMSE			Reports that overall ns but assume this is within group not between group comparison
Week 6	14.8 ± 4.0	15.7 ± 3.5	
Week 12	15.2 ± 3.8	16.2 ± 3.6	
Week 18	15.3 ± 4.1	16.3 ± 3.6	
Week 24	15.3 ± 4.3	16.3 ± 3.7	
Week 30	14.9 ± 4.4	16.0 ± 3.6	
PSMS			Reports that overall ns but assume this is within group not between group comparison
Week 6	9.1 ± 2.7	11.3 ± 4.1	
Week 12	9.1 ± 2.8	11.2 ± 3.9	
Week 18	9.0 ± 2.8	11.0 ± 3.7	
Week 24	9.0 ± 2.8	11.0 ± 3.7	
Week 30	9.1 ± 2.8	11.0 ± 3.7	

Comments: ADAS-cog measures cognition, the scale ranges from 0–70 where 70 = severe impairment
MMSE measures staging of disease and cognition, the scale ranges from 0–30 where 0 = severe impairment
PSMS measure ADLs, the scale ranges from 6–30 points, where 30 = severe impairment

Adverse effects (% occurrence)		
Nausea	8	15
Vomiting	5	10
Dizziness	10	15
Diarrhoea	8	10
Abdominal pain	5	8
Headache	8	10

Comments: neither treatment was associated with any significant adverse event; there was no significant difference in the incidence of adverse events in either of the groups. Most AEs were defined as mild to moderate and time-related. Nausea and vomiting occurred most frequently during the dose titration phase of rivastigmine, and no specific treatment was required. Dizziness, headache also occurred more frequently with higher dose of rivastigmine and resolved without treatment. No patient withdrew because of AEs or poor compliance to the treatment.

Methodological comments

- Allocation to treatment groups: patients were those consecutively referred to a neuropsychogeriatric ward. States that they were randomly assigned to treatment groups. No further details given.
- Blinding: not reported.
- Comparability of treatment groups: the treated groups were statistically similar in demographic and clinical characteristics.
- Method of data analysis: demographic data and general clinical features of both groups were analysed with the continuity-corrected χ^2 statistic and the Student's t-test when appropriate. Primary analyses for efficacy included repeated measures ANOVA for all rating scales used, with the treatment as a factor. This is a within-group comparison not a between-group comparison. Unclear how many participants were included in each of the evaluation points or at endpoint analysis.
- Sample size/power calculation: not reported.
- Attrition/dropout: not reported.

General comments

- Generalisability: mild-to-moderate probable AD.
- Outcome measures: appropriate.
- Inter-centre variability: single-centre study.
- Conflict of interests: not reported.

Quality criteria for Fuschillo *et al.*[68]

1. Was the assignment to the treatment groups really random?	Unknown
2. Was the treatment allocation concealed?	Unknown
3. Were the groups similar at baseline in terms of prognostic factors?	Reported
4. Were the eligibility criteria specified?	Adequate
5. Were outcome assessors blinded to the treatment allocation?	Inadequate
6. Was the care provider blinded?	Inadequate
7. Was the patient blinded?	Inadequate
8. Were the point estimates and measure of variability presented for the primary outcome measure?	Adequate
9. Did the analyses include an intention-to-treat analysis?	Inadequate
10. Were withdrawals and dropouts completely described?	Unknown

Reference and design	Intervention	Participants	Outcome measures
Author: Wilkinson *et al.*[69] Year: 2002 Country: International Study design: RCT Number of centres: 19 (UK, South Africa, Switzerland) Funding: sponsored by a grant from Eisai Inc and Pfizer Inc	Treatment arms: (dosing was in accordance with the respective approved product labelling at the time of the study) (1) donepezil 5 mg once daily for 28 days (the minimum interval specified in the product labelling), and then increased to 10 mg/day (one tablet) (2) rivastigmine 1.5 mg capsules twice daily (3 mg/day) with food initially. At 14-day intervals, the minimum interval specified in the labelling, patients were assessed and the dose increased to 3 mg (day 14), 4.5 mg (day 28) and finally to a maximum of 6mg twice daily (day 42), dependent upon tolerability The protocol allowed for dosing flexibility – the dose of either drug could be decreased to the previously administered dose level at	Number of participants: unclear how many screened. 112 patients randomised: 57 to Donepezil and 55 to Rivastigmine groups. 1 patient in the Donepezil group didn't receive any study medication, therefore results are based on 111 Sample attrition/dropout: those who were unable to tolerate the minimum effective doses, according to the product labelling (5 mg/day Donepezil, 6 mg/day Rivastigmine) were discontinued as per protocol. (number of patients remaining on treatment during the study noted below) Rates of discontinuation (%): Group 1) 6 (10.7) discontinued (AE related to study drug 4 (7.1), AE unrelated 2 (3.6): Group 2) 17 (30.9) discontinued (AE related to study drug 11 (20.0), AE unrelated 1 (1.8), death 1 (1.8), protocol violation 1 (1.8), didn't meet entrance criteria‡ 1 (1.8), withdrew consent/refusal to continue 2 (3.6). ‡ due to unreliable caregiver Sample crossovers: none reported Inclusion/exclusion criteria for study entry: patients at least 50 years of age with diagnostic evidence of mild to moderate, possible or probably AD consistent with DSM IV and NINCDS-ADRDA criteria.	Primary outcomes: ADAS-cog (11 item), MMSE Secondary outcomes: safety and tolerability: adverse events, physical examination, concomitant medication use, laboratory test abnormalities, vital signs, ECG, compliance. Adverse events and compliance data extracted where reported, but other outcome not as per review protocol. Ease of use and general satisfaction with dosing frequency and titration of the assigned treatment for each patient Methods of assessing outcomes: ADAS-cog was administered by independent raters blinded to study treatment and outcome; caregivers were instructed not to reveal assigned study medication. ADAS-cog is sensitive and reliable (see results for

continued

Reference and design	Intervention	Participants	Outcome measures
	any time during the study if the current dose wasn't tolerated. The dose could also be increased subsequently Other interventions used: to reflect clinical practice, dosage adjustments based on tolerability were permitted throughout the study Use of selective serotonin reuptake inhibitors, small daily doses of neuroleptics and short-acting benzodiazepines taken for insomnia and anxiety were permitted if given in stable doses for at least 1 month before study entry, and the frequency and dose remained unchanged. Most commonly taken medications were: Rheumatic diseases/gout: grp 1) 48.2%, grp 2) 36.4% antihypertensives: 1) 33.9%, 2) 21.8% Analgesics: 1) 17.9%, 2) 21.8% antibacterial: 1) 8.9%, 2) 25.5% sedatives/hypnotics/ anxiolytics: 1) 16.1%, 2) 12.7% antidepressants: 1) 7.1%, 2) 12.7% pyschoses/related disorders: 1) 1.8%, 2) 10.9%	MMSE of 10–26 inclusive, together with CT or MRI scan (within last 12 months) consistent with the diagnosis of AD. A caregiver, able to provide information on the patient's status and ensure compliance with treatment and clinic visits, was required. Those with stable and controlled concomitant diseases were included Patients treated previously with donepezil or rivastigmine were excluded, as were those taking medications with pronounced anticholinergic effects Characteristics of participants: (mean ± SD unless otherwise stated): Analysed for safety, AEs and lab data: 1) 56; 2) 55 Analysed for efficacy: 1) 56; 2) 54 (one did not have baseline assessment) Age in years (range): 1) 74.0 ± 7.6 (51–87); 2) 74.9 ± 7.3 (52–90) Months since diagnosis 1) 17.5 ± 19.3; 2) 19.3 ± 21.5 Female (%): 1) 30 (54); 2) 35 (64) Weight, kg M/F: 1) 76.8/59.2; 2) 74.1/60.0 MMSE (range): 1) 21.5 ± 4.1 (13–28); 2) 20.7 ± 4.9 (8–29) MMSE scores ≥ 21 (mild): 1) 38; 2) 34 MMSE scores ≤ 20 (moderate): 1) 18; 2) 21 Screening ADAS-cog (range): 1) 20.4 ± 7.8 (7–36); 2) 20.8 ± 8.8 (7–50) baseline ADAS-cog (range): 1) 20.2 ± 8.9 (8–41); 2) 20.6 ± 9.1 (9–47) presenting with ≥ 1 comorbid disease (%): 1) 42 (75.0); 2) 45 (81.8) taking ≥ 1 concomitant medication (%)†: 1) 48 85.7); 2) 50 (90.9) † Started before and continued during study or started after randomisation	scoring), and assesses selected areas of cognitive impairment (memory, language, orientation, reason and praxis). It was administered at screening, baseline and at weeks 4 and 12. The MMSE was undertaken at screening, baseline and weeks 4 and 12 by clinicians who had knowledge of the patients assigned medication Adverse events were assessed during clinic visits at weeks 4 and 12, and additionally at 2 and 6 for those receiving rivastigmine. In addition physicians performed telephone interviews with caregivers at weeks 6 (donepezil) and 8 (rivastigmine) for adverse events and compliance Satisfaction/ease of use by Likert-type questionnaires at weeks 4 and 12 to physicians and caregivers. Questionnaire developed by Pfizer and Eisai in conjunction with an external consultant (see below for scoring). Questions related to factors such as ease of use, satisfaction with dosing frequency, titration schedule and frequency of patient monitoring Length of follow-up: 12 weeks

continued

Results

Outcomes	**Donepezil (*n*'s below)**	**Rivastigmine (*n*'s below)**	
ADAS-cog, mean ±SE change from baseline			Treatment difference (95% CI)
Week 4	–2.14 ± 0.54 ($n = 53$)	–2.02 ± 0.61 ($n = 45$)	0.12 (–1.47, 1.71)
Week 12	–0.90 ± 0.56 ($n = 50$)	–1.05 ± 0.67 ($n = 37$)	–0.15 (–1.85, 1.55)
MMSE, mean ± SE change from baseline			
Week 4	1.09 ± 0.41 ($n = 53$)	0.16 ± 0.43 ($n = 50$)	–0.94 (–2.11, 0.23)
Week 12	0.71 ± 0.44 ($n = 51$)	1.20 ± 0.52 ($n = 39$)	0.49 (–0.82, 1.81)

Comments: both outcomes failed to show a difference between treatments at weeks 4 and 12. ADAS-cog ranges from 0 to 70 with higher scores indicating greater cognitive impairment. MMSE scores range from 0 to 30 with lower scores indicating greater cognitive impairment.

Satisfaction/ease of use (mean ± SE)	**Donepezil**	**Rivastigmine**	***p*-Value between groups**
Physicians			
Week 4	8.2 ± 0.4	12.9 ± 0.4	$p < 0.0001$
Week 12	8.5 ± 0.4	11.5 ± 0.5	$p < 0.0001$
Caregivers			
Week 4	9.9 ± 0.4	12.8 ± 0.5	$p < 0.0001$
Week 12	10.9 ± 0.5	12.4 ± 0.6	$p < 0.05$

Comments: Physicians satisfaction/ease of use questionnaire consisted of 6 questions (total score 6–30) and the caregiver's questionnaire had 8 questions (total score range 8–40) with lower scores indicating greater satisfaction/ease of use. Both physicians and caregivers reported better mean total scores for donepezil than rivastigmine.

Adverse effects

Treatment-emergent AEs occurring in ≥5% of patients and ≥ twice as frequently in either treatment group (all causalities)	**Donepezil *n* = 56** ***n* (%)**	**Rivastigmine *n* = 55** ***n* (%)**
Nausea	6 (10.7)	23 (41.8)
Vomiting	4 (7.1)	13 (23.6)
Headache	4 (7.1)	10 (18.2)
Anorexia	1 (1.8)	5 (9.1)
Abnormal dreams	4 (7.1)	1 (1.8)
Back pain	4 (7.1)	0
Somnolence	1 (1.8)	3 (5.5)
Urinary tract infection	3 (5.4)	0
Percent experiencing ≥ 1 treatment-related AE	42.9%	58.2%

Comments: the majority of treatment-related AEs in both groups were related to the digestive and nervous systems and were of mild or moderate severity. Four patients in the rivastigmine group reported a severe treatment-related digestive system AE (nausea, $n = 2$; vomiting, $n = 2$; abnormal liver function tests, $n = 1$) compared with none in the donepezil group

Withdrawal due to AE	6 (10.7%)	12 (21.8%)
Withdrawal possibly or definitely treatment related	4	11

continued

Number of patients remaining on treatment (estimated from figure except 12 weeks)			
Baseline	56	55	
Up to 4 weeks	55	51	
Up to 8 weeks	52	44	
Up to 12 weeks	50 (89.3%)	38 (69.1%)	$p = 0.009$
Reached maximum dose at some point during trial	98.2%	60.0%	
Remained at maximum dose until study completion/final visit	87.5%	47.3%	
Required a dose reduction or temporary discontinuation	17.9%	34.5%	
Dose adjustments throughout trial, *n* (%)	Always on 5 mg: 1 (1.8) 5–10 mg: 49 (87.5) 5–10–5 mg: 6 (10.7)	3 mg: 3 (5.5) 3–6 mg: 8 (14.5) 3–6–3–6 mg: 1 (1.8) 3–6–9 mg: 1 (1.8) 3–6–9–12 mg: 26 (47.3) 3–6–9–12–9 mg: 7 (12.7) 3–6–9–6 mg: 8 (14.5) 3–6–9–6–9–6 mg: 1 (1.8)	

Methodological comments

- Allocation to treatment groups: Patients assigned 1:1 to study groups using a stratified randomisation scheme based on centre and disease severity, i.e. mild (MMSE 21–26) and moderate (MMSE 10–20) dementia ensuring an equivalent distribution of patients with mild and moderate dementia between the treatment groups. No details of operationalisation of randomisation procedures or allocation concealment.
- Blinding: Open-label study. ADAS-cog was administered by independent raters blinded to study treatment and outcome. Other outcomes the assessor was not blinded.
- Comparability of treatment groups: Baseline patient demographic characteristics were similar for both treatment groups.
- Method of data analysis: In all cases ANCOVA was used as the primary model for estimating and testing treatment effects. Both OCs and LOCF/ITT analyses were performed for the population at weeks 4 and 12, although only the OCs are is presented because of the potential for bias resulting from large differences in study discontinuation rates between the two treatment groups. All two-sided, and at the 0.05 significance level. ITT does not meet criteria for ITT as one patient did not receive medication and was excluded, and at least one did not have any post-baseline assessments.
- Sample size/power calculation: Not described.
- Attrition/dropout: Rates reported and reasons given.

General comments

- Generalisability: Those over 50 years with mild to moderate dementia of the Alzheimer's disease type.
- Outcome measures: Cognitive measures appropriate. Unclear how reliable the satisfaction questionnaire is.
- Inter-centre variability: Not reported.
- Conflict of interests: Sponsored by a grant from Eisai Inc and Pfizer Inc. Rivastigmine is registered by Novartis who were not involved in the trial.

Quality criteria for Wilkinson *et al.*[69]

1. Was the assignment to the treatment groups really random?	Unknown
2. Was the treatment allocation concealed?	Unknown
3. Were the groups similar at baseline in terms of prognostic factors?	Reported
4. Were the eligibility criteria specified?	Adequate
5. Were outcome assessors blinded to the treatment allocation?	Partial
6. Was the care provider blinded?	Inadequate
7. Was the patient blinded?	Inadequate
8. Were the point estimates and measure of variability presented for the primary outcome measure?	Adequate
9. Did the analyses include an intention-to-treat analysis?	Inadequate
10. Were withdrawals and dropouts completely described?	Adequate

Reference and design	Intervention	Participants	Outcome measures
Author: Jones *et al.*[70] Year: 2004 Country: International Study design: RCT, multicentre Number of centres: 14 (UK, Finland, Germany, Norway) Funding: Eisai Inc and Pfizer Pharmaceuticals	Treatment arms: (1) Donepezil, administered orally and titrated to maximum effective therapeutic doses, according to product labelling: 5 mg once daily for 4 weeks, and then increased to 10 mg/day (2) Galantamine, administered orally and titrated to maximum effective therapeutic doses, according to product labelling: 4 mg twice daily (BD) for 4 weeks and then increased to 8 mg BD for a further 4 weeks, and then to 12 mg BD The protocol allowed flexibility in dosing such that if the current dose was not tolerated, the dose of either drug could be decreased to the previous dose at any time. Subsequently the	Number of participants: unclear how many were screened. Total randomised: 120: group 1) 64; group 2) 56. (Majority enrolled in UK and Finland centres, 46 (38.3%) and 44 (36.7%) respectively) Sample attrition/dropout: 61 (95.3%) of donepezil patients and 51 (91.1%) of galantamine patients completed the study. 3 (4.7%) donepezil and 4 (7.1%) galantamine patients discontinued due to adverse events. 1 subject default resulted in 1 discontinuation in the galantamine group Sample crossovers: none Inclusion/exclusion criteria for study entry: patients at least 50 years of age with probable or possible, mild or moderate AD consistent with DSM IV and NINCDS-ADRDA criteria. MMSE of 10–24 inclusive, together with CT or MRI scan (within last 18 months) consistent with the diagnosis of AD. A caregiver, able to provide information on the patient's status and ensure compliance with treatment and clinic visits, was required Patients treated previously with ChE inhibitors were excluded, as were patients with known hypersensitivity to ChE inhibitors, piperidine or alkaloid derivates	Primary outcomes: Physician's and Caregiver's Satisfaction Questionnaires (P&CSQ) Secondary outcomes: ADAS-cog (11-item and 13-item version) MMSE, Disability Assessment for Dementia (DAD) scale. Safety and tolerability: physical examination, vital signs, ECG, laboratory test abnormalities, concomitant medication use, compliance, and adverse events, Adverse events and compliance data extracted where reported, but other outcomes not as per review protocol Methods of assessing outcomes: P&CSQ rated ease of use and satisfaction with dosing frequency and titration of the assigned treatment at weeks 4, 8 and 12. (See below for scoring) Questionnaires were developed by Pfizer and Eisai in conjunction with an external consultant, and the study reports that use in a

continued

Reference and design	Intervention	Participants	Outcome measures
	dose could again be increased, with the intent to reach and maintain the highest recommended dose Other interventions used: none reported	and those who had been treated with any investigational drug within 30 days of the screening visit. Those taking medications with pronounced anticholinergic effects, such as drugs for Parkinson's disease, neuroleptics or tricyclic antidepressants within 1 month of study entry were also excluded. Those with significant obstructive pulmonary disease, asthma, gastrointestinal, endocrine or cardiovascular disease were not enrolled Characteristics of participants: (mean ± SD unless otherwise stated): Age, years (range): group 1) 73.8 ± 7.4 (51–88); group 2) 75.1 ± 7.7 (53–89) Age since onset of diagnosis (range): 1) 73.5 ± 7.5 (51–88); 2) 74.6 ± 7.9 (51–89) months since diagnosis, median (range): 1) 3.1 (0–47.5); 2) 3.2 (0–45.6) Female (%): 1) 33 (51.6); 2) 40 (71.4), $p = 0.03$ Screening MMSE (range): 1) 17.9 ± 3.3 (11–24); 2) 18.1 ± 3.2 (10–25) (see methodological comments below) No. with screening MMSE scores (%): mild (19–24): 1) 27 (42.1); 2) 26 (46.4) moderate (10–18): 1) 37 (57.8); 2) 27 (48.2) Baseline MMSE score (range): 1) 18.3 ± 3.3 (8–24); 2) 18.4 ± 3.7 (10–25) (see methodological comments below) ADAS-cog (11-item) (range): 1) 23.1 ± 7.4 (8–42); 2) 23.1 ± 8.7 (10–44) ADAS-cog (13-item) (range): 1) 32.5 ± 8.2 (16–54); 2) 32.8 ± 9.9 (17–56) DAD (range): 1) 25.9 ± 9.7 (3–40); 2) 25.4 ± 10.7 (0–40)	previous study supports the validity of the questionnaires Cognitive assessments were carried out at screening, baseline, and at weeks 4, 8, and 12 by independent raters who were blinded to the patients' assigned study medication and dosing regimen, and other study information such as AEs. Separate case report forms were used by these raters to maintain blinding of assigned treatment The 40-item DAD scale was completed by a trained rater/physician, with caregiver input, to assess both instrumental and basic ADL An AE was defined as any undesirable effect experienced by a patient during the trial, whether or not it was considered to be related to treatment. A serious AE was "life threatening or resulted in death, hospitalisation, prolongation of hospitalisation, or significant disability" Length of follow-up: 12 weeks

continued

Results

Outcomes	**Donepezil**	**Galantamine**	***p*-Value**
Cognition (all change from baseline and estimated from figures)			
ADAS-cog-11 (OC)			
Week 4	–2.2 $n = 61$	–0.7 $n = 54$	ns
Week 8	–2.9 $n = 60$	–2.1 $n = 53$	ns
Week 12	–4.9 $n = 60$	–2.5 $n = 52$	$p < 0.01$
ADAS-cog-11 (LOCF)			
Week 12	–4.7 $n = 64$	–2.3 $n = 55$	$p = 0.01$
% Degree of response ADAS-cog-11 at week 12			
≥7 points (substantial)	28.3	11.5	$p = 0.029$
≥4 (good)	53.3	28.8	$p = 0.009$
≥0	83.3	75.0	
MMSE (OC)			
Week 4	1.1 $n = 60$	0.2 $n = 54$	ns
Week 8	1.3 $n = 59$	1.4 $n = 49$	ns
Week 12	1.8 $n = 60$	1.2 $n = 48$	ns
MMSE (LOCF)			
Week 12	1.6 $n = 64$	0.8 $n = 56$	$p < 0.05$

Comments: The 13-item ADAS-cog has a score range of 0–85 and includes two items, delayed recall and concentration/distractability, in addition to those comprising the traditional, well-established 11-item ADAS-cog.
Text reports that ADAS-cog-13 also showed statistically significant improvement in the donepezil group compared with the galantamine group at 12 weeks (OC) and endpoint (LOCF), $p < 0.05$. Text also reports that the percent degree of response showed similar findings in the ADAS-cog-13 as the ADAS-cog-11 shown above.
Post-hoc analyses explored whether the 30% of galantamine patients who couldn't tolerate 24 mg/day influenced the observed difference in ADAS-cog scores between the groups. At week 12 (OC), stratified by dose of drug, patients with 10 mg/day donepezil showed a greater improvement on ADAS-cog-11 (LS mean treatment difference 2.49, $p < 0.05$) and the ADAS-cog-13 (LS mean treatment difference 2.78, $p < 0.05$) compared with galantamine patients receiving 24 mg/day.

Functional ability			
DAD (change from baseline & estimated from figure)			
Week 4	0.6 $n = 61$	–1.0 $n = 53$	$p < 0.01$
Week 8	1.1 $n = 60$	–0.2 $n = 53$	
Week 12	1.6 $n = 60$	–0.3 $n = 52$	$p < 0.05$
DAD week 12 (LOCF)	1.5 $n = 64$	–0.4 $n = 55$	$p < 0.05$

Comments: DAD is a 40-item scale and total scores range from 0–40, with lower scores indicating greater impairment of ADL. The DAD assessed each ADL on 3 aspects of executive function: initiation (consisting of the ability to decide and/or start an action); planning and organisation; and effective performance (consisting of the ability to complete an action).
The total DAD scores remained below baseline throughout the study in the galantamine group. At endpoint, the mean changes from baseline favoured donepezil compared with galantamine for 9 out of 10 individual items on the DAD. In the case of 2 (meal preparation, finance and correspondence), patients receiving donepezil showed significantly greater improvements compared to galantamine group ($p < 0.05$ and $p = 0.001$, respectively).

Satisfaction/ease of use (mean ± SE) (estimated from figure)	**Donepezil**	**Galantamine**	***p*-Value between groups**
Physicians (OC)			
Week 4	26.4 $n = 60$	25.6 $n = 54$	$p < 0.05$
Week 8	26.7 $n = 60$	25.9 $n = 53$	
Week 12	26.9 $n = 60$	25.0 $n = 52$	$p < 0.001$
Week 12 LOCF	26.7 $n = 64$	25.0 $n = 56$	$p < 0.001$
Caregivers (OC)			
Week 4	37.2 $n = 61$	36.2 $n = 54$	$p < 0.05$
Week 8	37.9 $n = 60$	37.3 $n = 53$	
Week 12	37.6 $n = 60$	35.8 $n = 51$	$p < 0.001$
Week 12 LOCF	37.2 $n = 64$	35.5 $n = 56$	$p < 0.001$

continued

Comments: Physicians satisfaction/ease of use questionnaire consisted of 6 questions (total score 6–30) and the caregiver's questionnaire had 8 questions (total score range 8–40) with higher scores indicating greater satisfaction. Responses to 4 of the 6 individual items of the physician's satisfaction questionnaire at endpoint favoured donepezil ($p < 0.01$) as follows: frequency of phone calls/visits concerning use of study medication and side effects, dosing frequency, titration schedule to clinically effective dose, and overall convenience of use. For the remaining 2 items (ease of use for patient and caregiver, overall satisfaction with medication), the responses favoured donepezil but were not statistically significant.
Responses to 3 of the 8 individual items (satisfaction with dosing frequency, tolerating the medication, physician contact about side effects) at endpoint showed statistical significance in favour of donepezil ($p < 0.05$). For the remaining 5 items the responses favoured donepezil but were not statistically significant.

Compliance

Maximum daily dose at some point in the study	98.4%	94.6%
Remained at max dose until study completion/final visit	92.2%	71.4%
Dose adjustments throughout trial, *n* (%)	Always on 5 mg: 1 (1.6) 5–10 mg: 58 (90.6) 5–10–5 mg: 4 (6.3) 5–10–5–10 mg: 1 (1.6)	Always on 8 mg: 1 (1.8) 8–16 mg: 2 (3.6) 8–16–24 mg: 40 (71.4) 8–16–24–16 mg: 13 (23.2)
Dose reduction due to AE	9.4%	21.4%

Comments: overall patients took the study drugs as prescribed by the clinician, and in general, patients in both groups took their daily doses as directed. Patients were permitted to return for unscheduled visits following clinic visits at weeks 4 and 8 in the event that a dose escalation was not tolerated. From weeks 4 to 8 there was no significant difference between the two groups regarding the number of patients requiring unscheduled visits (3 donepezil, 0 galantamine). From weeks 8 to 12 significantly more galantamine patients required unscheduled visits (0 donepezil, 13 galantamine, $p = 0.0001$). Overall, only 3 (4.7%) in the donepezil group had unscheduled visits due to dose tolerability problems compared with 13 (23.2%) in the galantamine group ($p < 0.01$).

Adverse effects

Treatment emergent AEs (all causalities)	43 (67.2%)	41 (73.2%)	
Treatment emergent AEs occurring in ≥ 5% of patients (all causalities)	*n* = 64	*n* = 56	
Nausea	10 (15.6)	13 (23.2)	* AEs occurring at twice the rate or more in the comparator group
Diarrhoea	6 (9.4)	8 (14.3)	
Anorexia	3 (4.7)	5 (8.9)	
Vomiting*	0	7 (12.5)	
Head ache	4 (6.3)	3 (5.4)	
Urinary tract infection*	2 (3.1)	4 (7.1)	
Dizziness*	1 (1.6)	3 (5.4)	
Withdrawn due to adverse events	3 (4.7%)	4 (7.1%)	
Cause of withdrawal (adverse event related to study drug)	1 (depression)	3 (depression, vomiting, nausea)	

Comments: more than 95% of AEs experienced were mild to moderate in nature. Serious AEs were experienced by 4 (6.3%) donepezil patients (congestive heart failure, abscess, skin disorder, atrial fibrillation and hyponatraemia), and 2 (3.6%) galantamine patients (heart failure and hypotension). However, none of these were considered by the investigator to be treatment related.

Methodological comments

- Allocation to treatment groups: Patients were randomised via a computer-generated schedule to receive either drug according to a 1:1 ratio. Patients were further stratified based on disease severity at screening, e.g. mild (MMSE 19–24) and moderate (MMSE 10–18) dementia.
- Blinding: Open-label study. Independent raters who were blinded to the patients' assigned study medication and dosing regimen assessed cognitive outcomes.
- Comparability of treatment groups: Treatment groups were similar for age and disease severity. Gender distribution differed between groups (51.6% in group 1 were women compared to 71.4% in group 2, $p = 0.03$). One patient had a screening MMSE of 25 (baseline 23) in group 2 and 1 a score of 25 at baseline (screening 23), and 1 patient in group 1 had a baseline MMSE of 8 (screening 11).

continued

- Method of data analysis: For the P&CSQs at endpoint used an LOCF approach for an ITT population. ANOVA was used to estimate and test treatment differences. Observed case (OC) analyses were also performed for the ITT population at each visit (weeks 4, 8, 12). In addition, treatment groups were compared for individual items of the P&CSQ using Cochran–Mantel–Haenszel test. For secondary efficacy assessments ANCOVA models were used with terms for treatment, centre/country and baseline (weeks 4, 8, and 12 OC, and week 12 LOCF). Response rates on the 11-item and 13-item ADAS-cog scores were compared with Cochran–Mantel–Haenszel test. Adverse event data were summarised and no statistical tests were conducted. All statistical tests were two-sided, and with 0.05 significance level. It would appear that there is one less patient in the LOCF group for galantamine, therefore doesn't meet criteria for ITT.
- Sample size/power calculation: Determination of study size based on total scores of P&CSQs. To have an 80% chance of detecting a difference of two points between treatment groups on both parameters, with an overall Type I error of 5%, 60 patients in each group were required.
- Attrition/dropout: Numbers and reasons for withdrawals reported.

General comments
- Generalisability: Those aged over 50 years with mild to moderate Alzheimer's disease type dementia.
- Outcome measures: Cognitive measures appropriate. Unclear how reliable the satisfaction questionnaire is, and the DAD not reported.
- Inter-centre variability: Not reported.
- Conflict of interests: Sponsored by Eisai and Pfizer Inc. and two authors are employed by these companies. Galantamine is licensed by Shire who were not involved in the study.

Quality criteria for Jones *et al.*[70]

1. Was the assignment to the treatment groups really random?	Adequate
2. Was the treatment allocation concealed?	Unknown
3. Were the groups similar at baseline in terms of prognostic factors?	Reported
4. Were the eligibility criteria specified?	Adequate
5. Were outcome assessors blinded to the treatment allocation?	Partial
6. Was the care provider blinded?	Inadequate
7. Was the patient blinded?	Inadequate
8. Were the point estimates and measure of variability presented for the primary outcome measure?	Inadequate
9. Did the analyses include an intention-to-treat analysis?	Inadequate
10. Were withdrawals and dropouts completely described?	Adequate

Appendix 11

Data extraction: memantine RCTs

Reference and design	Intervention	Participants	Outcome measures
Author: Reisberg *et al*.[72] Year: 2003 Country: USA Study design: Randomised, placebo-controlled, double-blind, parallel group study Number of centres: 32 Funding: Merz Pharmaceuticals (Frankfurt, Germany) and National Institute on Aging of the National Institutes of Health	Treatment arms: (1) memantine 20 mg per day vs (2) placebo (identical in appearance) 28-week study, mean ± SD duration of treatment for both groups 24 ± 8 weeks Other interventions used: Patients could be receiving antidepressant treatment providing this had been stable for at least 2 months. Chloral hydrate was also permitted (as a sedative or hypnotic) but could not be used 24 hours before an assessment	Number of participants: 345 patients were screened and of these 252 were randomly assigned to study groups. Memantine = 126 Placebo = 126 Sample attrition/dropout: 71 patients discontinued treatment before week 28, memantine *n* = 29, placebo *n* = 42 Sample crossovers: There was no crossover in this study Inclusion/exclusion criteria for study entry: At least 50 years of age, residing in the community. Probably AD according to DSM-IV and NINCDS-ADRDA. MMSE: 3 to 14, GDS stage 5 or 6, FAST stage 6a or greater, signifying the presence of dementia-related deficits in the ability to perform one or more basic activities of daily living Reliable caregivers; must have undergone CT or MRI of the brain within the previous 12 months Excluded Patients with vascular dementia, dementia or clinically significant neurologic disease due to conditions other than AD, major depressive disorder, score greater than 4 on the modified Hachinski Ischemic Rating Scale. Patients with clinically significant coexisting medical conditions or laboratory abnormalities. Patients receiving specific concomitant medications (anticonvulsant agents, antiparkinsonian agents, hypnotic agents, anxiolytic agents, neuroleptic agents, cholinomimetic agents, or any other investigational compounds) Characteristics of participants: ± values are mean ± SD Memantine *n* = 126 / Placebo *n* = 126 / Total *n* = 252	Primary outcomes: CIBIC-plus global score at 28 weeks (measures cognition, function and behaviour). Alzheimer's Disease Cooperative Study-Activities of Daily Living modified for severe dementia (ADCS/ADLsev) change from baseline to week 28. A subgroup of 19 individually validated items (the ADCS/ADLsev) was used (measures ability to perform items of the activities of daily living, ranging from total independence to total inability) Secondary outcomes: Severe Impairment Battery (SIB) (evaluates cognitive performance). MMSE (evaluates cognitive function). Global Deterioration Scale (GDS) (assesses overall cognitive and functional capacity). Functional Assessment Staging Scale (FAST) (assesses magnitude of progressive functional deterioration in patients with dementia by identifying characteristic progressive disabilities). Neuropsychiatric Inventory (NPI) (assesses neuropsychiatric disturbances). Resource Utilisation in Dementia instrument (to assess the burden on the caregiver, also to provide Alzheimer's disease-related health economics data) Assessments conducted at baseline, week 12 and week 28 (end of treatment), or at early termination with a 28 week retrieved-dropout visit when possible (Only 5 of the 71). If 28-week observation values were not available the last observed value was used

continued

Reference and design	Intervention	Participants	Outcome measures

	Finished *n* = 97	Withdrew *n* = 29	Finished *n* = 84	Withdrew *n* = 42	
Female (%)	70 (72.2)	21 (72.4)	55 (65.5)	24 (57.1)	170 (67.5)
Male (%)	27 (27.8)	8 (27.6)	29 (34.5)	18 (42.9)	82 (32.5)
Age – year	75.5 ± 8.16	77.3 ± 9.17	75.8 ± 7.28	77.5 ± 8.61	76.1 ± 8.07
Education – year	12.3 ± 3.06	13.0 ± 3.14	12.9 ± 3.14	11.7 ± 2.91	12.5 ± 3.09
Race (%)					
White	85 (87.6)	27 (93.1)	75 (89.3)	40 (95.2)	227 (90.1)
Black	4 (4.1)	1 (3.4)	4 (4.8)	2 (4.8)	11 (4.4)
Other	8 (8.2)	1 (3.4)	5 (6.0)	0	14 (5.6)
MMSE score	7.8 ± 3.76	7.6 ± 3.67	8.1 ± 3.60	7.9 ± 3.54	7.9 ± 3.64
GDS stage (%)					
5	46 (47.4)	13 (44.8)	41 (48.8)	12 (28.6)	112 (44.4)
6	51 (52.6)	16 (55.2)	43 (51.2)	30 (71.4)	140 (55.6)

Measures of safety: assessed at specified intervals, included neurologic and physical examinations, measurement of vital signs, electrocardiography, laboratory tests and recording of adverse events

Methods of assessing outcomes:
CIBIC-plus: Included domains for cognition assessed by patient interview; function assessed by caregiver interview; behaviour assessed by separate interviews of caregiver and patient. Interviews conducted by experienced clinicians blinded to adverse events and other study assessments. The same clinician completed all CIBIC-plus interviews for each study patient and associated caregiver wherever possible.
ADCS/ADLsev: caregivers assessed a patient's activities during the preceding 4-week interval. The differences in total scores were analysed.
GDS: assessed by observations of the patient and reports from the caregiver.
NPI: scored based on information from the caregiver regarding the patient's behaviour and associated distress felt by the caregiver.
Resource Utilisation in Dementia instrument, through structured interviews with caregivers

To assess the clinical relevance of treatment effects further, a multifactor responder analysis was predefined

Length of follow-up: 28 weeks but as mean ± SD duration of treatment for both groups was 24 ± 8 weeks it seems that some participants were treated for longer than 28 weeks

Data analysis was performed by a contract research organisation (Quintiles) along with the authors

continued

Results

Outcomes	**Memantine (*n* = 126 at start)**	**Placebo (*n* = 126 at start)**	***p*-Value**
Deaths	2	5	
Comments: Two patients (one in each group) died in the 30-day period after the last dose of study medication. All deaths were considered to be unrelated to study medication.			
CIBIC-plus	Mean ± SD	Mean ± SD	
Change from baseline at endpoint, LOCF analysis **	4.5 ± 1.12 (*n* = 118)	4.8 ± 1.09 (*n* = 118)	0.06
Change from baseline at week 28, analysis of observed cases	4.4 ± 1.12 (*n* = 97)	4.7 ± 1.13 (*n* = 84)	0.03
CIBIC-plus Global Score Values in *italics* estimated from graph	Mean ± SE	Mean ± SE	
Baseline	4.0 (*n* = 126)	4.0 (*n* = 126)	
Week 12	*4.23 ± 0.10* (n = *107*)	*4.30 ± 0.10* (n = *105*)	
Week 28	*4.39 ± 0.11* (n = *97*)	*4.74 ± 0.11* (n = *84*)	0.03 (mean difference 0.3; 95% CI –0.69 to –0.03)
Endpoint	*4.66 ± 0.10* (n = *118*)	*4.79 ± 0.10* (n = *118*)	0.06 (mean difference 0.3; 95% CI –0.51 to 0.02)
Comments: scored on a seven-point scale ranging from 1 (markedly improved) to 7 (markedly worse). **Sixteen patients were excluded from this analysis because they had not been assessed after the baseline assessment. *p*-Values are based on the Wilcoxon–Mann–Whitney test for between-treatment comparisons. The CIBIC-plus is a change score. By design the base-line score, "no change", is set at 4.00. Higher values at subsequent measurements indicate worsening. The endpoint and week 28 values are actual mean ratings.			
ADCS/ADLsev	Mean ± SD	Mean ± SD	
Change from baseline at endpoint, LOCF analysis **	–3.1 ± 6.79 (*n* = 124)	–5.2 ± 6.33 (*n* = 123)	0.02
Change from baseline at week 28, analysis of observed cases	–2.5 ± 6.27 (*n* = 97)	–5.9 ± 6.78 (*n* = 84)	0.003
Difference in ADCS/ADLsev Score Values in *italics* estimated from graph	Mean ± SD	Mean ± SD	
baseline	0 (*n* = 126)	0 (*n* = 126)	
Week 4	*0.4 ± 0.3* (n = *119*)	*–0.6 ± 0.3* (n = *117*)	
Week 12	*–0.6 ± 0.6* (n = *107*)	*–2.1 ± 0.5* (n = *106*)	
Week 28	*–2.4 ± 0.7* (n = *97*)	*–5.9 ± 0.7* (n = *84*)	0.003 (mean difference 3.4; 95% CI 1.45; 5.28)
Endpoint (LOCF)	*–3.1 ± 0.6* (n = *124*)	*–5.2 ± 0.6* (n = *123*)	0.02 (mean difference 2.1; 95% CI 0.49; 3.78)
Comments: A total score of 54 signified optimal performance, and lower scores indicated worse performance. **Five patients were excluded from this analysis because they had not been assessed after the baseline assessment. Baseline scores were 26.8 (memantine group, *n* = 126) and 27.4 (placebo group, *n* = 126)			
SIB	Mean ± SD	Mean ± SD	
Change from baseline at end point, LOCF analysis	–4.0 ± 11.34 (*n* = 124)	–10.1 ± 13.50 (*n* = 123)	<0.001
Change from baseline at week 28, analysis of observed cases	–4.5 ± 11.48 (*n* = 96)	–10.2 ± 12.66 (*n* = 83)	0.002
Difference in SIB score Values in *italics* estimated from graph	Mean ± SD	Mean ± SD	
baseline	0 (*n* = 126)	0 (*n* = 126)	
Week 4	*1.0 ± 0.8* (n = *119*)	*–0.4 ± 0.8* (n = *117*)	
Week 12	*0.8 ± 1.0* (n = *107*)	*–5.4 ± 1.2* (n = *106*)	
Week 28	*–4.4 ± 1.2* (n = *96*)	*–10.2 ± 1.4* (n = *83*)	0.002
Endpoint±(LOCF)	*–4.0 ± 1.0* (n = *124*)	*–10.2 ± 1.2* (n = *123*)	<0.001

continued

Comments: a 51-item scale, assesses social interaction, memory, language, visuo-spatial ability, attention, praxis and construction. Scored from 0 (greatest impairment) to 100. Baseline scores were 65.9 (memantine group, $n = 126$) and 68.3 (placebo group, $n = 126$). On the basis of the predetermined definition of a response in the study protocol, 29% of the patients receiving memantine and 10% of those receiving placebo had a response ($p < 0.001$). Values in italics estimated from graph

MMSE	Mean ± SD	Mean ± SD	
Change from baseline at end-point, LOCF analysis	–0.5 ± 2.40 ($n = 124$)	–1.2 ± 3.02 ($n = 124$)	0.18 (NS)
Change from baseline at week 28, analysis of observed cases	–0.6 ± 2.61 ($n = 97$)	–0.9 ± 3.09 ($n = 82$)	0.68 (NS)

Comments: A 30-point scale, higher scores indicate better function. Baseline scores were 7.7 (memantine group, $n = 126$) and 8.1 (placebo group, $n = 126$) (also see Characteristics of Participants table above)

GDS	Mean ± SD		
Change from baseline at end-point, LOCF analysis	0.1 ± 0.47 ($n = 121$)	0.2 ± 0.48 ($n = 119$)	0.11 (NS)
Change from baseline at week 28, analysis of observed cases	0.1 ± 0.49 ($n = 97$)	0.2 ± 0.48 ($n = 84$)	0.16 (NS)

Comments: A seven-stage scale, higher stages signify greater impairment. Baseline scores were 5.5 (memantine group, $n = 126$) and 5.6 (placebo group, $n = 126$)

FAST	Mean ± SD		
Change from baseline at end-point, LOCF analysis	–0.2 ± 1.24 ($n = 121$)	0.6 ± 1.39 ($n = 118$)	0.02
Change from baseline at week 28, analysis of observed cases	0.1 ± 1.24 ($n = 97$)	0.5 ± 1.38 ($n = 84$)	0.007
Difference in FAST score	Mean ± SD	Mean ± SD	
Values in *italics* estimated from graph			
baseline	0 ($n = 126$)	0 ($n = 126$)	
Week 12	*0 ± 0.09* (n = *107)*	*0.09 ± 0.10* (n = *106)*	
Week 28	*0.10 ± 0.13* (n = *97)*	*0.51 ± 0.14* (n = *84)*	0.007
Endpoint	*0.19 ± 0.11* (n = *121)*	*0.57 ± 0.13* (n = *118)*	0.02

Comments: Seven major stages range from stage 1 (normal) to stage 7 (severe dementia). Five substages in stage 6 correspond to the loss of ability to independently dress, bath and handle proper mechanics and cleanliness in using the toilet, and to remain continent. Six substages in stage 7 correspond to loss of speech, ambulation and other motor capacities. The FAST scores were calculated by enumerating the FAST stages and substages as follows: stage 3 (a score of –2) through 5 (0) and substage 6a (1) through substage 7*f* (11). Baseline scores were 2.8 (memantine group, $n = 126$) and 2.8 (placebo group, $n = 126$)

NI	Mean ± SD		
Change from baseline at endpoint, LOCF analysis	0.5 ± 15.76 ($n = 120$)	3.8 ± 16.06 ($n = 119$)	0.33 (NS)
Change from baseline at week 28, analysis of observed cases	0.1 ± 15.92 ($n = 97$)	2.9 ± 16.13 ($n = 84$)	0.60 (NS)

Comments: A 12-item scale with scores ranging from 0 to 144 for the patient assessment rating and from 0 to 60 for the caregiver-distress rating, with 0 indicating the optimum in both cases. These NPI scores are from the patient assessments. Baseline scores were 21.4 (memantine group, $n = 126$) and 19.5 (placebo group, $n = 126$)

Resource Utilisation in Dementia Score	Indicated caregivers spent less time with patients receiving memantine	
LOCF analysis	Difference between groups 45.8 hours (95% CI 10.37 to 81.27)	0.01

General comments: Additional analyses were performed with different strategies used for missing values, as described in the protocol. The results were unchanged in each of these analyses (results not shown in paper).
A subgroup analysis examined whether efficacy was seen in both patients with moderate AD (MMSE 10–14) and severe AD (MMSE <10). A benefit or memantine as compared with placebo was suggested for all outcome measures in both groups (results not shown in paper).

continued

Withdrawals *n* (%) due to		
adverse events	13 (10%)	22 (17%)
withdrawal of consent	12 (10%)	14 (11%)
death of participant	1 (1%)	4 (3%)
protocol violation	3 (2%)	3 (2%)
change of caregiver	0	2 (2%)

Comments: Patients could have multiple reasons for discontinuation. Of those discontinuing due to adverse events, agitation was the most common reason for discontinuing (5% of memantine, and 7% of placebo)

Adverse effects *n* (%)		
Any adverse event	106 (84)	109 (87)
Agitation	23 (18)	40 (32)
Urinary incontinence	14 (11)	14 (11)
Urinary tract infection	7 (6)	17 (13)
Insomnia	13 (10)	10 (8)
Diarrhoea	12 (10)	10 (8)

Comments: All patients were included in the safety analysis. The above are the most frequently reported AEs occurring in at least 10% of the patients in either treatment group. Most AEs were mild to moderate in severity and were either not related or unlikely to be related to the study medication. The incidence rates for the frequently reported adverse events in the memantine group were no more than 2% higher than in the placebo group. Serious AEs were reported in 16 (13%) patients receiving memantine and 23 (18%) patients receiving placebo. Most of the serious AEs were considered to be unrelated to study medication. Also see Deaths, at top of this section.

Methodological comments

- Allocation to treatment groups: Randomisation was stratified according to site with the use of RanCode (version 3.1) and in blocks of four. Staff at the individual sites were blinded (unclear how) to the randomisation process.
- Blinding: Placebo was identical in appearance to study medication but no other details are given. Clinicians conducting interviews for the primary outcome CIBIC-plus were blinded to adverse events and other study assessments. The same clinician completed all CIBIC-plus interviews for each study patient and associated caregiver wherever possible. It is not made clear whether caregivers were blinded to treatment group or not.
- Comparability of treatment groups: Reports that baseline characteristics were similar, but no statistical comparison given. The way the baseline data are presented makes it hard to judge comparability between groups. Participants in each group have been split into study completers and non-study completers. The memantine group seems to have slightly more women (72%) overall than the placebo group (63%).
- Method of data analysis: The main efficacy analysis was based on the randomised patients who received at least one assessment after baseline (some patients withdrew before assessment and were excluded; the authors describe this as ITT but this may not include all randomised participants). This analysis included both those who completed the study and those who discontinued their assigned treatment prematurely. For the latter, the efficacy observation at week 28 was imputed from the last available observation carried forward. Three additional analyses were also performed to adjust for missing values. One analysis was identical to the primary analysis, except that the actual retrieved-dropout values at week 28 were used when available. A second analysis included patients for whom no value after baseline was available in addition to the "intention-to-treat population" and assumed no change in the outcome measures for these patients (ITT in the reviewer's opinion). In the third analysis, missing values were replaced for those with no value after base line by the mean observed value for decline in the placebo group (also ITT). An observed-cases analysis was also performed based on data for all randomised patients who were available for evaluation at week 28. Only the primary analysis (LOCF on the author-defined ITT population) and the observed case analysis were presented in the paper. Means and either SD or SE were reported 95% CIs were given in the text for some outcomes. Wilcoxon-Mann-Whitney test used.
- Sample size/power calculation: There is no mention of how the study authors decided how many participants to recruit.
- Attrition/dropout: Assigned treatment was discontinued if continuation represented a medical risk in the opinion of the study physician, or if the patient declined ongoing participation. Subjects who withdrew prematurely were asked to complete endpoint measures at the time of early termination and to return at 28 weeks for a "retrieved-dropout visit", which included all endpoint assessments. Only 5 of the 71 patients who left the study returned for a retrieved-dropout visit. Reasons for participants dropping out are reported.

General comments

- Generalisability: The results would not necessarily be applicable to people with moderate to severe AD in addition to other clinically significant conditions because such people were excluded from this study.
- Outcome measures: Outcome measures appear to be relevant to this study area and seem to have been measured appropriately.
- Inter-centre variability: It is not possible to judge whether there was any between-centre variability. All clinicians were described as experienced, so hopefully this would help to reduce variability. Primary outcomes were assessed by the same clinician.
- Conflict of interests: Merz Pharmaceuticals provided study medication and funding and was involved in planning the design and protocol. Three authors have received lecture fees from Lundbeck, one has also received lecture fees from Forest and Merz. Two authors received consulting fees from Forest and Lundbeck. One author received consulting fees from Forest.

Quality criteria for Reisberg *et al.*[72]

1. Was the assignment to the treatment groups really random?	Adequate
2. Was the treatment allocation concealed?	Unclear
3. Were the groups similar at baseline in terms of prognostic factors?	Reported
4. Were the eligibility criteria specified?	Adequate
5. Were outcome assessors blinded to the treatment allocation?	Partial
6. Was the care provider blinded?	Adequate
7. Was the patient blinded?	Adequate
8. Were the point estimates and measure of variability presented for the primary outcome measure?	Adequate
9. Did the analyses include an intention-to-treat analysis?	Inadequate
10. Were withdrawals and dropouts completely described?	Adequate

Reference and design	Intervention	Participants	Outcome measures
Author: Tariot *et al.*[71] Year: 2004 Country: USA Study design: Double-blind RCT, multicentre Number of centres: 37 Funding: Forest Research Institute (a division of Forest Laboratories Inc).	Treatment arms: (1) Memantine 20 mg/day (2) Placebo Treatment was titrated in 5-mg weekly increments from a starting dose of 5 mg/day to the target of 20 mg/day at the beginning of week 4. From week 3 to the end of week 8 of double-blind treatment, transient dosage adjustments were permitted for patients experiencing dose-limiting adverse events. All patients receiving memantine were required to receive the target dose of 20 mg/day by the end of week 8. Patients not tolerating the target dose by week 8 were disenrolled	Number of participants: 589 patients assessed for eligibility of whom 404 were randomised to two groups: Gp 1: Memantine = 203 (reduced to 202 as 1 withdrew consent before receiving treatment) Gp 2: Placebo = 201 Sample attrition/dropout: Memantine gp: a further 30 discontinued study participation during the trial. Placebo gp: 51 discontinued study participation during the trial. No patients were excluded during the placebo lead-in period for lack of compliance. No patients discontinued due to changes in administration of donepezil Sample crossovers: None Inclusion/exclusion criteria for study entry: Diagnosis of probable AD (NINCDS-ADRDA). MMSE score of 5 to 14 at both screening and baseline; minimum age 50 years; a recent (within 12 months) MRI or CT scan consistent with diagnosis of probable AD; ongoing cholinesterase inhibitor therapy with donepezil for more than 6 months before entrance to trial and at a stable dose (5–10 mg/day) for at least 3 months; a knowledgeable and responsible caregiver to accompany the patient to research visits and oversee the administration of the investigational agent	Primary outcomes: SIB – change from baseline and a modified ADCS/ADL at week 24 Secondary outcomes: Clinician's Interview-Based Impression of Change Plus Caregiver Input (CIBIC-plus); NPI; and the BGP Methods of assessing outcomes: SIB: a 40-item test developed for the evaluation of cognitive dysfunction in patients with more severe AD. Six primary subscales assess memory, orientation, language, attention, visuo-spatial ability and construction. In addition, the scale assesses praxis, social interaction, and orienting to name. Validity, reliability and sensitivity to longitudinal change have been established. Scores range from 0 to 100 (higher scores indicate higher levels of cognitive ability) ADCS/ADL: A 19-item subset of the original 42-item inventory focusing on items appropriate for the assessment of later stages of dementia (i.e. the level of independence in everyday tasks such as eating, walking, grooming, telephone use, hobbies, complex

continued

Reference and design	Intervention	Participants	Outcome measures
	Other interventions used: All patients were to maintain stable donepezil therapy at entry dose as prescribed by the patient's physician for the duration of the study. Stable doses of concomitant medications permitted included antidepressants, antihypertensives, anti-inflammatory drugs, atypical antipsychotics, antiparkinsonian drugs, anticoagulants, laxatives, diuretics, and sedatives/ hypnotics. Adherence to this protocol was monitored by routine assessment of concomitant medication use Adherence with study medication was assessed by returned tablets and more than 95% of both treatment groups had more than 75% compliance (95% for placebo group and 96.5% for memantine group)	during the trial; residence in the community; ambulatory or ambulatory-aided (i.e. walker or cane) ability; and stable medical condition. Exclusion criteria: clinically significant B_{12} or folate deficiency; active pulmonary, gastrointestinal, renal, hepatic, endocrine or cardiovascular disease; other psychiatric or central nervous system disorders; CT or MRI evidence of clinically significant central nervous system disorders other than probable AD; dementia complicated by other organic disease; or a modified Hachinski Ischaemia Score of more than 4 at screening. Patients were discontinued from the study if the inclusion criterion of concomitant donepezil therapy was no longer met Characteristics of participants: Baseline characteristics provided for those who received study medication so Gp1: Memantine $n = 202$ & Gp2: placebo $n = 201$ Data are mean (SD) unless stated Men, *n* (%): Gp1 74 (37); Gp2 67 (33). Women, *n* (%): Gp1 128 (63); Gp2 134 (67). Age, year: Gp1 75.5 (8.45); Gp2 75.5 (8.73). Weight, kg: Gp1 70.7 (14.31); Gp2 66.4 (14.12). White race *n* (%): Gp1 182 (90.1); Gp2 186 (92.5). MMSE score: Gp1 9.9 (3.13); Gp2 10.2 (2.98). Duration of donepezil treatment, week: Gp1 126 (64.9); Gp2 129 (70.3). Donepezil dose, mg: Gp1 9.25 (1.79); Gp2 9.49 (1.88). Any concurrent medical condition, *n* (%): Gp1 149 (73.8); Gp2 149 (74.1). Any concomitant medication during treatment, *n* (%): Gp1 197 (97.5); Gp2 197 (98.0). Tocopherol, *n* (%): Gp1 131 (64.9); Gp2 120 (59.7). Multivitamins, *n* (%): Gp1 80 (39.6); Gp2 78 (38.8). Acetylsalicylic acid, *n* (%): Gp1 73 (36.1); Gp2 76 (37.8). Ascorbic acid, *n* (%): Gp1 43 (21.3); Gp2 35 (17.4). Paracetamol, *n* (%): Gp1 32 (15.8); Gp2 25 (12.4). Ginkgo biloba, *n* (%): Gp1 31 (15.3); Gp2 24 (11.9). Calcium, *n* (%): Gp1 25(12.4); Gp2 21 (10.4) Participants in the memantine group were slightly heavier ($p = 0.003$) than those in the placebo group. Most patients (87%) had a FAST rating between 4 and 6c. The most frequent medication classes (>20%) used during treatment were: Vitamins Gp 1 77%, Gp2 74%. Analgesics Gp 1 48%, Gp2 48%. Antidepressants Gp 1 36%, Gp2 36%. Mineral	tasks and communications). The sensitivity and reliability of this modification have been established. Scores range from 0 to 54, higher scores indicating higher functioning. Administered as an interview to the patient's caregiver CIBIC-plus: administered according to the format of the AD CS-CGIC. Assesses the effect of medication on overall clinical status in patients with dementia, incorporating caregiver observations as well as patient interviews. Change rated on a scale from 1 (marked improvement) to 7 (marked deterioration). A global assessment of severity of illness was made at baseline NPI: to assess the frequency and severity of behavioural symptoms in patients with dementia, based on an interview of the caregiver. The 12-item version of the instrument was used with a total score ranging from 0 to 144. Higher scores reflect greater symptoms. Assessed at baseline, at the end of week 12 and at the final visit only BGP: 35 items (scored 0, 1 or 2 by the rater) assessing observable aspects of cognition, function and behaviour. A higher score reflects worse function. The BGP care dependency subscale reflects cognitive and functional characteristics associated with increased need for care. The BGP was administered at baseline and the final visit only FAST was administered as an index of staging, and not as a secondary outcome, at baseline and the final visit Concomitant medications and vital signs were recorded at every visit; adverse events were recorded at baseline and all subsequent visits Laboratory tests, electro-cardiograms and physical examinations were performed at the screening and final visits (data not extracted as per protocol)

continued

Reference and design	Intervention	Participants	Outcome measures
		supplements Gp 1 27%, Gp2 22%. Lipid-reducing agents Gp 1 25%, Gp2 23%. Anxiolytics/neuroleptics Gp 1 22%, Gp2 26%. Anti-inflammatory agents Gp 1 24%, Gp2 21% The most frequent medical conditions were not recorded, but the following body systems were noted to be affected at screening: eyes-ears-nose-throat: Gp 1 43%, Gp2 43%; neurological Gp 1 38%, Gp2 34%; appearance/skin Gp 1 33%, Gp2 40%; musculokeletal Gp 1 29%, Gp2 29%; cardiovascular Gp 1 23%, Gp2 20%; abdomen Gp 1 17%, Gp2 12%; head/neck Gp 1 9%, Gp2 6%; other Gp 1 9%, Gp2 10%; pulmonary Gp 1 5%, Gp2 3% Baseline scores for outcome measures [mean (SD)] SIB Gp1 77.8 (15.46); Gp 2 79.8 (14.18) ADCS/ADL Gp1 35.9 (9.75); Gp 2 36.2 (9.32)	Length of follow-up: 24 weeks (following a 1–2 week single-blind placebo lead-in before randomisation to assess compliance) Outcome measures were obtained at baseline and at the end of weeks 4, 8, 12, 18 and 24. Patients who discontinued prematurely were evaluated during their final visit

Results Statistically significant benefits of treatment with memantine vs treatment with placebo were observed on all primary and secondary outcome measures as presented.

Outcomes	Memantine (n = 203)	Placebo (n = 201)	p-Value
Completed study	172 (85.1%)	150 (74.6%)	0.01
SIB [least-squares mean score (SE)] Baseline	78.0 (1.11) n = 198	80.0 (1.13) n = 197	
Changes from baseline	0.9 (0.67) n = 198	–2.5 (0.69) n = 196	<0.001
End point LOCF	1.0 (0.70) n = 171	–2.4 (0.74) n = 153	<0.001
Week 24 Observed cases			
LS Mean difference between mean change from baseline values			
Week 4 (observed case)	–1.2 (Gp 1 n = 197; Gp 2 n = 194)		0.06
Week 8 (observed case)	–1.5 (Gp 1 n = 190; Gp 2 n = 180)		0.03
Week 12 (observed case)	–3.1 (Gp 1 n = 185; Gp 2 n = 169)		<0.001
Week 18 (observed case)	–2.7 (Gp 1 n = 181; Gp 2 n = 164)		0.006
Week 24 (observed case)	–3.4 (Gp 1 n = 171; Gp 2 n = 153)		<0.001
Endpoint (LOCF)	–3.4 (Gp 1 n = 198; Gp 2 n = 196)		<0.001
Comments: Range of possible scores 0 to 100, higher score indicates better function (higher level of cognitive ability). Analyses using the LOCF approach showed a statistically significant benefit of memantine treatment vs treatment with placebo on the SIB (p < 0.001) as did analyses using the OC approach (p < 0.001). Post-hoc analyses including all randomised patients also showed statistically significant benefits consistent with analyses using the modified intent-to-treat population (p < 0.001). Only patients with at least 1 postbaseline assessment were included in the LOCF analysis. The endpoint is the last non-missing postbaseline assessment carried forward to end of study. The mean change from baseline by visit and at endpoint using OC and LOCF, showed statistically significant differences between the memantine and placebo groups at all visits beginning at week 8. The mean SIB values for the patients receiving memantine remained above baseline throughout the trial.			
ADCS/ADL$_{19}$ [LS mean score (SE)] Baseline	35.5 (0.73) n = 198	35.8 (0.74) n = 197	
Changes from baseline	–2.0 (0.50) n = 198	–3.4 (0.51) n = 197	0.03
Endpoint LOCF	–1.7 (0.51) n = 172	–3.3 (0.55) n = 152	0.02
Week 24 Observed cases			

continued

LS Mean difference between mean change from baseline values			
Week 4 (observed case)	–0.8 (Gp 1 $n = 198$; Gp 2 $n = 195$)		0.03
Week 8 (observed case)	–1.1 (Gp 1 $n = 190$; Gp 2 $n = 182$)		0.01
Week 12 (observed case)	–1.3 (Gp 1 $n = 185$; Gp 2 $n = 170$)		0.02
Week 18 (observed case)	–1.4 (Gp 1 $n = 181$; Gp 2 $n = 163$)		0.03
Week 24 (observed case)	–1.6 (Gp 1 $n = 172$; Gp 2 $n = 152$)		0.02
Endpoint (LOCF)	–1.4 (Gp 1 $n = 198$; Gp 2 $n = 197$)		0.03

Comments: Range of possible scores 0 to 54, higher score indicates better function. Analyses using the LOCF approach showed a statistically significant benefit of memantine treatment vs treatment with placebo on the ADCS/ADL_{19} ($p < 0.03$) as did analyses using the observed case approach ($p < 0.02$). Post-hoc analyses including all randomised patients also showed statistically significant benefits consistent with analyses using the modified intent-to-treat population ($p < 0.03$). The mean change in total ADCS/ADL from baseline by visit and at end point by using OC and LOCF showed a statistically significant difference ($p < 0.05$) from placebo beginning at week 4.

CIBIC-plus [arithmetic mean (SE)]	NA $n = 198$	NA $n = 197$	
Baseline			
Changes from baseline	4.41 (0.074) $n = 198$	4.66 (0.075) $n = 196$	0.03
Endpoint LOCF	4.38 (0.081) $n = 172$	4.64 (0.087) $n = 152$	0.03
Week 24 Observed cases			

Distribution (% of patients) of CIBIC-plus ratings at endpoint (LOCF) *values in italics estimated from figure*	Memantine ($n = 198$)	Placebo ($n = 196$)	
Marked improvement	*1*	*0*	$p = 0.03$ for the
Moderate improvement	*4*	*2*	comparison
Minimal improvement	*8*	*10*	between the
No change	*4*	*33*	distribution of
Minimal worsening	*31*	*32*	values for the
Moderate worsening	*14*	*20*	memantine and
Marked worsening	*1*	*3*	placebo groups

Comments: Defined as a change score, therefore baseline values are not applicable. Range of possible scores 1 (marked improvement) to 7 (marked worsening). The mean CIBIC-plus score was statistically significantly better for the memantine groups vs the placebo group using both OC and LOCF. 55% of memantine gp rated as improved or unchanged vs 45% of placebo gp at endpoint.

NPI [LS mean score (SE)]	13.4 (1.07) $n = 198$	13.4 (1.08) $n = 197$	
Baseline			
Changes from baseline	–0.1 (0.98) $n = 193$	3.7 (0.99) $n = 189$	0.002
Endpoint LOCF	–0.5 (0.99) $n = 171$	2.9 (1.06) $n = 152$	0.01
Week 24 Observed cases			

Range of possible scores 1 to 144, higher scores indicate worse symptoms. The total NPI score was significantly lower for the memantine group compared with the placebo group at week 24, representing fewer behavioural disturbances and psychiatric symptoms for patients in the memantine group

BGP Care Dependency Subscale			
[LS mean score (SE)]	9.5 (0.45) $n = 198$	9.8 (0.46) $n = 196$*	
Baseline			
Changes from baseline	0.8 (0.37) $n = 185$	2.3 (0.38) $n = 179$	0.001
Endpoint LOCF	0.6 (0.37) $n = 172$	2.2 (0.40) $n = 151$	0.001
Week 24 Observed cases			

*One patient had an incomplete BGP baseline assessment and was not included.

Range of possible scores 0 to 70, higher scores indicate worse function. The BGP care dependency subscale was statistically significantly improved for the memantine group compared with the placebo group

Withdrawals before receiving study medication	1	0

continued

Withdrawals after receiving study medication	30	51	
Adverse events	15 (7.4%)	25 (12.4%)	
Withdrew consent	8	16	
Protocol violation	1	5	
Insufficient therapeutic response	1	3	
Other	5	2	

Comments: More participants in the placebo-treated group discontinued prematurely because of adverse events than in the memantine group. The adverse event most often associated with discontinuation was confusion resulting in discontinuation in 1.5% of patients in the placebo group and 2% in the memantine group.

AEs reported in at least 5% of patients in either treatment group	78% data from $n = 202$ patients	72% data from $n = 201$ patients	
Agitation	19 (9.4%)	24 (11.9%)	
Confusion	16 (7.9%)	4 (2.0%)	$p = 0.1$
Fall	15 (7.4%)	14 (7.0%)	
Influenza-like symptoms	15 (7.4%)	13 (6.5%)	
Dizziness	14 (6.9%)	16 (8.0%)	
Headache	13 (6.4%)	5 (2.5%)	$p = 0.9$
Urinary tract infection	12 (5.9%)	10 (5.0%)	
Urinary incontinence	11 (5.4%)	6 (3.0%)	
Accidental injury	10 (5.0%)	16 (8.0%)	
Upper respiratory tract infection	10 (5.0%)	13 (6.5%)	
Peripheral edema	10 (5.0%)	8 (4.0%)	
Diarrhoea	9 (4.5%)	17 (8.5%)	
Fecal incontinence	4 (2.0%)	10 (5.0%)	

Comments. Patients may have reported more than one adverse event. Most adverse events were rated as mild or moderate in severity and were judged to be not related to study drug for participants in both treatment groups. The only adverse events that occurred in at least 5% of the memantine group and with an incidence of at least twice that of the placebo group were confusion and headache. By similar criteria lower incidences of diarrhoea and faecal incontinence were observed in the memantine group compared with the placebo group. Other gastrointestinal effects of interest for patients receiving cholinesterase inhibitors included nausea (memantine gp 0.5%; placebo gp 3.5%) and constipation (memantine gp 3.0%; placebo gp 1.5%). Four (25%) of 16 patients experiencing confusion discontinued because of this adverse event compared to 3 (75%) of 4 patients receiving placebo who experienced this adverse event. In most of the patients receiving memantine, confusion was rated as mild, occurred at a median of 32 days and remitted within 2 weeks. In patients receiving placebo, confusion was more likely to be rated as severe, occurred at a median of 55 days and did not remit. No patients discontinued because of headache, which usually lasted 1 day.

Methodological comments

- Allocation to treatment groups: Patients allocated to treatment groups in permuted blocks of 4 in accordance with the randomisation list generated and retained by the Department of Biostatistics at Forest Laboratories. At the baseline visit, each investigator sequentially assigned a randomisation number to each patient. No individual patient randomisation code was revealed during the trial.
- Blinding: Masked study medication was supplied to each study site for dispensation in blister packs at each visit. Drug and placebo tablets were visually identical and all patients received 4 tablets of study medication daily (in combinations of memantine [5 mg] and matching placebo tablets).
- Comparability of treatment groups: Participants in the memantine group were slightly heavier ($p = 0.003$) than those in the placebo group. The authors retrospectively added this variable to the analysis of covariance and it did not affect the outcomes. No clinically relevant group differences were observed for the duration of donepezil use before baseline or for any other characteristic at baseline. Most patients (87%) had a FAST rating between 4 and 6c. There were no statistically significant differences between groups in the number or type of medical disorders experienced previously or at the time of enrolment, or in the number or type of concomitant medications used during the study.
- Method of data analysis: Three populations were considered. The randomised population consisted of all patients randomised into the study ($n = 404$); the safety population consisted of all randomised patients who received at least 1 dose of double-blind study medication ($n = 403$); the modified ITT population specified by the protocol consisted of patients in the safety population who completed at least 1 postbaseline SIB or ADC-AD_{19} assessment ($n = 395$). The statistical analysis plan for this study stipulated that only postbaseline data could be carried forward. Particularly for the CIBIC-plus, it is not possible to carry forward baseline data because by definition this is a change score

continued

and is not applicable to baseline. All efficacy analyses were based on the modified ITT population. Primary efficacy analyses were conducted by using the LOCF approach for missing data imputation (with only postbaseline assessments carried forward). Supportive analyses were performed by using the OC approach. Change from baseline was compared between groups using a two-way analysis of covariance, with the treatment group and centre as main effects and baseline total score as the covariate. The study was to be declared positive if memantine was statistically significantly better than placebo ($p < 0.5$) on both the SIB and ADCS-AD_{19}. For categorical measures, the Cochran–Mantel–Haenszel statistic using modified Ridit scores (Van Elteren test) [no further explanation or references given] controlling for study centre was used to compare distributions between memantine and placebo groups. No interim analyses were performed.

- Sample size/power calculation: Assuming a hypothetical effect size of 0.35, a sample size of at least 170 patients in each treatment group provided a 90% power at a two-sided α level of 0.05, based on a two-sample *t* test for change from baseline to week 24 in both SIB and ADCS/ADL_{19} scores. No rationale for picking a hypothetical effect size of 0.35 was given.
- Attrition/dropout: One patient dropped out from the memantine group before receiving study medication. Once on study medication 30 patients discontinued from the memantine group and 51 discontinued from the placebo group. No patients discontinued because of lack of compliance (during the placebo lead-in period) or because of changes in donepezil administration. Reasons for participants dropping out during the study and numbers dropping out are given

General comments

- Generalisability: Patients with moderate to severe AD who are already taking donepezil. A number of concomitant medications were permitted but anyone with additional clinically significant disease was excluded from this study and therefore the results may not be applicable to this group. Results may not be applicable to patients who are not already taking donepezil. The trial did not address different doses or titration rates, the use of other cholinesterase inhibitors besides donepezil or the impact of commencing memantine therapy before donepezil.
- Outcome measures: Outcome measures appear to be relevant to this study area and seem to have measured appropriately.
- Inter-centre variability: None reported. No information regarding procedures that may have been in place to limit variability is given.
- Conflict of interests: Forest Laboratories provided all financial and material support for the research, consulted with the authors and the members of the memantine study group on the study design, monitored the conduct of the study as well as the collection of the data, analysed and interpreted the data, and assisted the authors in the preparation, review and approval of the manuscript. Three of the authors are employees with stock options at Forest Laboratories.

Quality criteria for Tariot et *al.*[71]

1. Was the assignment to the treatment groups really random?	Adequate
2. Was the treatment allocation concealed?	Adequate
3. Were the groups similar at baseline in terms of prognostic factors?	Reported
4. Were the eligibility criteria specified?	Adequate
5. Were outcome assessors blinded to the treatment allocation?	Unclear
6. Was the care provider blinded?	Adequate
7. Was the patient blinded?	Adequate
8. Were the point estimates and measure of variability presented for the primary outcome measure?	Adequate
9. Did the analyses include an intention-to-treat analysis?	Inadequate
10. Were withdrawals and dropouts completely described?	Adequate

Appendix 12

Critique of submissions from industry: clinical effectiveness

SHTAC review of the industry submission for assessment of donepezil in patients with mild to moderately-severe Alzheimer's disease: clinical effectiveness

The submission appears to be a non-systematic review of the evidence. No details of any search strategy for identifying the trials are given. It is not clear how primary studies were assessed for inclusion, and eligibility criteria are not reported. No assessment of study quality has been performed.

The 2004 submission reports on nine studies which were available since the original appraisal in 2000. Five of these were included in the SHTAC review, although one of these was only included in the SHTAC review in part as the full trial included patients with severe AD. One trial included in the submission was previously only available as an abstract but was included in the submission as commercial in confidence (full paper to be submitted for publication) and this also met the inclusion criteria for the SHTAC review. The remaining three studies were excluded from the SHTAC review as the participant characteristics did not meet the SHTAC criteria. One was in patients with Downs Syndrome and dementia, and two were in patients with dementia associated with cerebrovascular disease.

Three additional RCTs published since 2001 and included in the SHTAC review were not detailed in the submission.

The submission also details eight non-randomised observational studies that have been completed since the submission in 2000. These studies were not controlled clinical trials and therefore did not meet the inclusion criteria for the SHTAC review.

The Eisai submission for the use of donepezil, which was an update, concluded that there is a substantial amount of new evidence that continues to support the demonstrable benefits of treatment with donepezil. These benefits can be realised in multiple domains including cognition, global function, ADL function, behaviour and quality of life. The submission also reports that early cessation of donepezil treatment may result in re-emergence or worsening of AD symptoms.

SHTAC review of the industry submission for assessment of rivastigmine in patients with mild to moderately severe Alzheimer's disease: clinical effectiveness

The industry submission appears to be a non-systematic review of evidence for the clinical efficacy of rivastigmine that has become available since the original submission in July 2000. The report does not provide any inclusion/exclusion criteria, and there is no information included on any search strategy used. It is therefore not possible to determine whether all relevant research was identified, and it is not clear how primary studies were assessed for inclusion. The submission does not include any assessment of study quality.

The submission includes 31 studies published since 2000, of which eight had already been excluded by SHTAC (two open label trials, three retrospective analyses and three other studies which were not RCTs). The remaining studies included six open label studies, 11 retrospective studies, two trials which included the wrong patient groups (one moderately-severe to severe AD, and one on dementia with Lewy Bodies) and two other studies which were not controlled clinical trials and therefore did not meet the inclusion criteria for the SHTAC review. The remaining two studies were highlighted as commercial in confidence, but these were not included in full in the submission, so could not be assessed for inclusion/exclusion by SHTAC.

Four additional RCTs published since 2001 and included in the SHTAC review were not detailed in the submission.

The Novartis industry submission for the use of rivastigmine concluded that new evidence and an updated health economic model reaffirm the clinical and cost-effectiveness of rivastigmine for the treatment of mild to moderately severe AD, defined by an MMSE score of 10–26. Other observations included a delay or deceleration in cognitive decline in patients treated for up to 5 years and pronounced benefits on cognitive performance and ADLs in rapidly progressing patients. Novartis also found that rivastigmine is effective in about half of patients who switch to rivastigmine following lack/loss of response to donepezil.

SHTAC review of the industry submission for assessment of galantamine in patients with mild to moderately severe Alzheimer's disease: clinical effectiveness

The submission appears to be a non-systematic review of evidence. No details of any search strategy for identifying the trials are given. It is not clear how primary studies were assessed for inclusion, and eligibility criteria are not reported. However, the submission appears to be an update of that made in the previous Alzheimer's disease review in 2000, and so more details may have previously been provided.

The 2004 submission reports on four studies which were available since the original appraisal in 2000. One of these studies had already been included in the SHTAC review and one was subsequently included as a result of it having been highlighted in the industry submission. The remaining two studies were excluded from the SHTAC review due to them including the wrong patient groups (both studies included patients with AD combined with cerebrovascular disease, and one also included patients with vascular dementia).

All but one of the other RCTs (Wilkinson, 2000[66]) included in the SHTAC review were reported in the submission. This study however, was a conference proceeding and this may therefore explain its exclusion from the industry submission. As well as clinical trials the industry submission also discusses subsequent open-label extension studies, which have not been included in the SHTAC review due to them not meeting the inclusion criteria.

The Shire industry submission for the use of galantamine concluded that data from six placebo-controlled trials show treatment benefits for patients across cognition, global clinical assessment, behaviour and ADL. This translates into reduced burden for informal caregivers. There is no apparent decline in treatment effect with disease severity. The submission also reports that patients who may be excluded by current guidance demonstrate benefits, and evidence shows that stopping treatment in such patients results in rapid symptomatic decline.

SHTAC review of the industry submission for assessment of memantine in patients with moderately severe to severe Alzheimer's disease: clinical effectiveness

The submission appears to be a non-systematic review of the evidence. No details of any search strategy for identifying the trials are given. It is not clear how primary studies were assessed for inclusion, and eligibility criteria are not reported. No assessment of study quality has been performed.

The submission reports the two studies included in the SHTAC review, namely those by Reisberg and colleagues[72] and Tariot and colleagues.[63] One further study is reported by Winblad and Poritis which had previously been excluded from the SHTAC review on the basis that it included a large proportion of patients with vascular dementia. One paper included in the submission which had not previously been identified for the SHTAC review was retrieved and assessed for inclusion/exclusion. The study has since been excluded on the basis that it was a retrospective study of previous trial data.

A subgroup analysis is discussed in the submission which investigated the use of memantine in patients with moderately severe AD. However, this was not a planned comparison, the number of participants was small, and it did not meet the SHTAC review inclusion criteria. One prospective, open, uncontrolled study is also discussed.

No additional studies were identified within the submission that met the inclusion criteria of the SHTAC review.

The Lundbeck industry submission concludes that several trials have reported the clinical

effectiveness of memantine in patients with moderately severe to severe AD. These data show the benefits in global function, cognition, ADL and behaviour with a resultant delay in time to total dependency and need for full-time care. Clinical trial data have shown that memantine is associated with a reduction in the number of hours of caregiver support (a reduction of 51.5 hours per month compared with placebo), a reduced risk of institutionalisation, and functional benefits in fundamental aspects of patients' self-care and independence, including use of the toilet, ability to drink, eat, dress, move and stand up. These clinical benefits of memantine have been demonstrated in treatment-naïve patients and patients withdrawn from acetylcholinesterase inhibitor (AChEI) therapy, as well as in patients receiving stable doses of an AChEI (donepezil).

Appendix 13

Critical appraisal of published cost-effectiveness studies (by product)

Economic evaluations of donepezil

Internal validity of donepezil studies

Note ? means unclear or unknown.

✓ means item included or judged as acceptable to be internally valid.

× means factor not included or judged as unacceptable to be internally valid.

Item/study	Stein 1997[81]	Stewart *et al.* 1998[82]	Jönsson *et al.* 1999[84]	O'Brien *et al.* 1999[85]	Neumann *et al.* 1999[86]
1. Well-defined question	✓	✓	✓	✓	✓
2. Clear description alternatives	✓ Donepezil 5 mg, and 10 mg, versus placebo	✓ Donepezil 5 mg, and 10 mg, versus placebo	✓ Donepezil 5 mg, and 10 mg, versus placebo	✓ Donepezil 5 mg versus no treatment	✓ Donepezil (5 mg or 10 mg) versus placebo
3. Reasonable study type	? CUA – by simple calculation	✓ CEA model	✓ CEA model	✓ CEA model	✓ CEA model (CUA)
4. Effectiveness established	✓ Cites phase II trial reported by Rogers 1996.[53] Also uses the limited information (unpublished at the time) from a 2-year observational extension of the Rogers trial and two phase III trials provided by Eisai Ltd. Presents summary findings	× Cites trial by Rogers 1998[51] but does not present any findings from this trial	✓/? Cites trial by Rogers 1998[51] and presents a short narrative summary (no numerical results) of the findings	✓ Cites trial by Rogers 1998[51] and presents a short narrative summary, including some numerical results, of the findings	× Cites trial by Rogers 1998[51] but does not present any findings from this trial

continued

Item/study	Stein 1997[81]	Stewart et al. 1998[82]	Jönsson et al. 1999[84]	O'Brien et al. 1999[85]	Neumann et al. 1999[86]
5. Effectiveness estimates related to population risks	× Trial population data from Rogers et al. 1996[53] (USA patients). No discussion of how the trial patient characteristics relate to the UK population	×/? Trial population data from Rogers et al. 1998[51] (USA patients) used to estimate disease progression in hypothetical cohorts of patients with mild or moderate AD. Cohort data from elderly people in Cambridge[111] (UK) used to estimate disease progression in severe AD subgroup. Two sources of mortality rates[112,113,129] (location of populations not reported). Authors acknowledge trial data comes from a different set of people but state it can be assumed that they are very similar to the UK population (no evidence is given to support this)	×/? The two models used effectiveness data from Rogers et al. 1998[51] (USA patients). Disease progression either calculated from the Kungsholmen Project195 (Swedish population) or the Rogers trial data.[51] Mortality rates taken from a cohort study[194] (Swedish population). Authors acknowledge that the trial patient characteristics differ from the Swedish study population which included patients with all types of dementia, not only AD. Patients in the clinical study were younger than in the Swedish study	?/✓ Trial population data from Rogers et al. 1998[51] (USA patients) used to estimate disease progression in a hypothetical cohort of persons with mild and moderate AD. Mortality rates and weighting factors come from a clinic-based AD cohort in Alberta, Canada.[114] No discussion of how the trial patient characteristics relate to the Canadian population	✓ Trial population data from Rogers et al. 1998[51] (USA patients) used to estimate disease progression in a hypothetical cohort of persons with mild and moderate AD. Disease progression estimated using data from the Consortium to Establish a Registry for Alzheimer's Disease[180] (CERAD) (USA population)

continued

Item/study	Stein 1997[81]	Stewart et *al.* 1998[82]	Jönsson et *al.* 1999[84]	O'Brien et *al.* 1999[85]	Neumann et *al.* 1999[86]
6. Relevant costs and consequences identified	*Costs:* ✓/? Drug costs used in the analysis. Direct costs for outpatient and memory clinic visits, and CT scanning estimated but not used in the analysis. Not included: some formal care costs (residential care, social services) and all informal care costs. *Consequences:* ?/× Estimate QALY gains (method unclear)	*Costs:* ? Cites papers reporting costs of typical care packages according to levels of cognitive disability.[118–120] Costs contributing to these care packages not summarised. Drug acquisition costs presented separately. *Consequences:* ? Estimates years spent in a non-severe AD state	*Costs:* ✓/? Cites the paper in which cost calculations are thoroughly described[193]. Accommodation, home help, and additional medication costs are included. No further details of included costs are provided. Consequences: ? Model estimates years spent with mild to moderate dementia	*Costs:* ✓ Cites resource use from the Canadian Study of Health and Aging (CSHA)[195] used to calculate long term care, community services, medications and unpaid caregiver time costs. Physician visit costs (for monitoring of care) also included. Not included: costs of acute care services (states these are the same for AD and non-AD persons of the same age) *Consequences:* ? Model estimates years spent with mild to moderate dementia	*Costs:* ✓ Cites the paper from which cost data were derived for direct medical and non-medical costs plus the costs associated with unpaid caregiving[196]. Additionally drug costs and costs associated with extra clinical visits (for monitoring drug efficacy, adequacy of dose and presence of side effects) were estimated and included. *Consequences:* ? Model estimates QALY gains

continued

Item/study	Stein 1997[81]	Stewart *et al.* 1998[82]	Jönsson *et al.* 1999[84]	O'Brien *et al.* 1999[85]	Neumann *et al.* 1999[86]
7. Costs and consequences measured accurately	*Costs:* ✓/? Use of clinic visits and CT scanning services estimated. No information given to support the estimates of service use. *Consequences:* ?/X QALYs calculated using three assumptions: 1) cognitive decline measures disease progression, 2) mortality unaffected by donepezil treatment, and 3) after 6 months donepezil has no effect on rate of disease progression. Possible errors in the measurement of consequences	*Costs:* ? Resource use data for cost estimates were from a study of elderly people with dementia with costs of typical care packages estimated for elderly people at different levels of cognitive disability. The proportion of these people with AD or other dementias is not stated and no information is given to indicate that costs for AD would be the same as for other dementias *Consequences:* ? Two illness progression scales used (1) MMSE in the cohort study; (2) OPCS SEVINT in the costs of care packages. Authors cite paper describing methodology used to transform SEVINT scores onto the MMSE scale. Model assumes drug treatment affects disease progression only for the first six months of the 5-year time horizon. Annual survival rates were calculated from a 3-year survival rate. Possible errors in the measurement and calculation of consequences	*Costs:* ? No patients with a diagnosis of dementia in the Kungsholmen database had MMSE scores of 27–30, therefore the mean costs for non-demented patients in that state was used. *Consequences:* ? Half-year transition probabilities for the model were calculated from data collected at intervals a few years apart. Treatment effects were taken into account as a risk reduction factor applied to the transition probabilities for the Kungsholmen population over the 5-year time horizon. Mortality data added to the Within-Trial analysis came from the general population. Possible errors in the measurement and calculation of consequences	*Costs:* ? Before estimating costs authors adjusted CSHA data for the oversampling of persons in institutions. *Consequences:* ? Transition state probabilities for the two groups are derived from trial data. Model assumes that after the 24-week treatment effect the benefits are maintained but not increased for the remainder of the 5-year time horizon. Possible errors in the measurement and calculation of consequences	*Costs:* ? Costs for moderate AD were assumed to lie midway between the costs of mild and severe AD. The number of extra office visits needed by patients on donepezil was based on discussions with clinicians. *Consequences:* ? Transition probabilities were obtained from CERAD data using a modified survival analysis. Treatment effects were taken into account as a risk reduction factor. Because these were similar for the 5 mg, 10 mg and pooled 5 mg and 10 mg groups the pooled 5 mg and 10 mg results were used. The base-case 18-month expected duration of drug effect was the mean from a survey of 32 clinical experts. The assumption that 4% of patients would discontinue treatment every six weeks after the first 6 months of treatment was based on a long term open label study. Possible errors in the measurement of consequences

continued

Item/study	Stein 1997[81]	Stewart *et al.* 1998[82]	Jönsson *et al.* 1999[84]	O'Brien *et al.* 1999[85]	Neumann *et al.* 1999[86]
8. Costs and consequences valued credibly	*Costs:* ✓ UK cost sources used. Source of drug costs not provided, unclear which year these relate to. Costs of outpatient and memory clinic visits, and CT scan calculated as the average of extra contractual referrals prices in 1996/7 (this data not included in final analysis). *Consequences:* × QALYs gained from the use of donepezil estimated using the IHQL (not validated for valuing cognitive impairment). No rationale given for the IHQL values used	*Costs:* ? Sources of costs used in papers cited, not summarised by authors. Source of drug costs not provided. *Consequences:* ? Consequences calculated from the economic model, over a 5-year time horizon, in terms of years spent in a non-severe AD state	*Costs:* ? Sources of cost information are provided but not summarised by authors. Drug costs are for 1998 but it is not clear which year the non-drug costs relate to. Cumulative 5-year cost of care is calculated from the economic model. *Consequences:* Consequences calculated from the economic model based on the number of patient years in non-severe AD states over 5 years	*Costs:* ✓ Nursing home costs from reimbursement information in Ontario. Medication costs (other than donepezil) taken from Ontario Drug Benefit Plan; a 10% pharmacy mark-up and a dispensing fee was added. Community services costs estimated from Hamilton region data on prices and volumes of services. Caregiver time valued at the minimum wage rate for Ontario. Resource use costs calculated from CSHA data. The economic model calculated expected cost per patient over 5 years *Consequences:* ? Consequences calculated from the economic model, over a 5-year time horizon, in terms of expected time in different AD states and expected years per patient in a non-severe AD state	*Costs:* ? Source of cost information is provided but not summarised by authors *Consequences:* ? Calculated from the model in terms of change in costs per change in QALY over 18 months. A threshold analysis (at 24 months) was conducted by modelling
9. Differential timing considered	✓ 6% discount rate for costs.	✓ 6% discount rate for costs.	✓ 3% discount rate for costs. Not applied to all outcomes.	✓ 5% discount rate for costs.	✓ 3% discount rate for costs and QALYs.
10. Incremental analysis performed	✓	✓	✓	✓	✓
11. Sensitivity analysis performed	✓ Two-way	✓ One-way	✓ One-way	✓ One-way	✓ One-way
12. Modelling conducted reasonably	Simple decision analysis model	Markov model used	Markov model used	Markov model used	Markov model used

Donepezil – continued

Item/study	Ikeda *et al.* 2002[87]	Sobolewski *et al.* 2002[88] (Abstract)	Fagnani *et al.* 2003[89]	Wimo *et al.* 2003[56]	AD2000 Collaborative Group 2004[43]
1. Well-defined question	✓	✓	✓	✓	✓
2. Clear description alternatives	✓ Donepezil 5 mg versus usual care	✓ Donepezil 5 mg, and 10 mg, versus no treatment	✓ Donepezil 10 mg versus no treatment	✓ Donepezil 5 mg, and 10 mg versus placebo	✓ Donepezil 5 mg, and 10 mg versus placebo
3. Reasonable study type	✓ CEA model (CUA)	? CEA model	✓ CMA model	? Cost-consequence analysis	? Cost-consequence analysis
4. Effectiveness established	× Cites Homma *et al.*[44] but no clear statement on drug efficacy given. Trial not summarised	× No studies cited, no information given	× Cites Winblad *et al.*[47] but does not present any findings from this trial	✓ Cites Winblad *et al.*[47] and presents a short narrative summary, including some numerical results, of the findings	✓ Paper reports clinical outcomes and concludes that any benefits with donepezil are below minimally relevant thresholds.
5. Effectiveness estimates related to population risks	?/× Trial population data from Homma *et al.*[44] (Japanese study), risk ratios applied to Neumann's transition probability data[86] (US population). Authors recognise that USA epidemiological data may not relate to Japanese population	? Not able to judge as abstract contains insufficient information	?/✓ Trial population data from Windblad *et al.*[47] (Five northern European countries: Denmark, Finland, The Netherlands, Norway and Sweden). Other data from French sources. No discussion of possible differences between the populations	✓ Individual patient data from the trial are used to calculate the cost differences between the two trial groups	✓ Individual patient data from the trial are used to calculate the cost differences between the two trial groups.

continued

Item/study	Ikeda *et al.* 2002[87]	Sobolewski *et al.* 2002[88] (Abstract)	Fagnani *et al.* 2003[89]	Wimo *et al.* 2003[56]	AD2000 Collaborative Group 2004[43]
6. Relevant costs and consequences identified	*Costs:* ✓ Based on insurance system values. Medical costs include outpatient visits. Long term care includes both institutional and home care. Not included: drug costs (other than donepezil) *Consequences:* ? QALY gains	*Costs:* ✓/? Direct costs only including medications, medical equipment, home help and community nurse. *Consequences:* ? Years in a state less than severe	*Costs:* ✓ Cites other studies that reported direct and indirect costs including caregiver time, hostel costs, medical expenditures, paid and unpaid care. *Consequences:* ? Proportion of non-dependent patients and residential status of patients	*Costs:* ✓ Direct and indirect costs including study drug, hospitalisation (patients and caregivers), visits to health care professionals (patients and caregivers), concomitant medications (patients and caregivers), living accommodation (patients only), social services (patients only), caregiver time. *Consequences:* ? Delay in loss of activities of daily living and cost savings	*Costs:* ✓ Costs included visits by health care professionals, domestic help, meals on wheels, day care, overnight stays in hospital, and overnight stays in nursing or residential homes. Reported costs did not include those of donepezil or institutionalisation. *Consequences:* ? Entry to institutional care, progression of disability and cost savings

continued

Item/study	Ikeda *et al.* 2002[87]	Sobolewski *et al.* 2002[88] (Abstract)	Fagnani *et al.* 2003[89]	Wimo *et al.* 2003[56]	AD2000 Collaborative Group 2004[43]
7. Costs and consequences measured accurately	*Costs:* ? Based on the long-term care insurance entitlement. Eligibility levels for different severities of AD estimated following consultation with experts. Assumed donepezil-treated patients visited outpatients twice as often as untreated patients (based on prescribing regulations). Outpatient costs based on hospitals of 200 beds or more. Dispensing costs based on an independent pharmacy handling over 4000 prescriptions per month. Drug costs based on 100% compliance *Consequences:* ? The transition probabilities calculated by Neumann *et al.*[86] were used. Treatment effects from the Homma trial[44] were taken into account as risk reduction factors. The duration of drug effect was presumed to last for the whole 2-year time horizon in the base-case analysis. Possible errors in the measurement of consequences	*Costs:* ? Cohort split 50:50 into mild and moderate disease states with 50% of patients on 5 mg and 50% on 10 mg donepezil. Rationale for these proportions not given *Consequences:* ? Assumes drug treatment duration and effects for only the first 6 months of the 5-year time horizon. Possible errors in the measurement of consequences	*Costs:* ? For patients living in the community net costs were used, measured by deducting the costs of elderly non-AD persons living at home from those of the AD population *Consequences:* ? Disease progression in the first year was based on the results of the cited clinical trial. Disease progression for the remainder of the 3-year time horizon was based on the natural course of cognitive decline in untreated AD patients according to the Stern growth model (referenced). Possible errors in the measurement of consequences	*Costs:* ✓ Determined from results of Resource Utilisation in Dementia questionnaire completed by caregivers of all trial participants and an average cost per patient calculated *Consequences:* ✓ No modelling required as no extension beyond the 1-year trial period	Costs: ✓ Determined from a survey completed by carers of formal services used by patients and carers and an average cost per patient calculated *Consequences:* ✓ No modelling required as no extension beyond the trial period

continued

Item/study	Ikeda *et al.* 2002[87]	Sobolewski *et al.* 2002[88] (Abstract)	Fagnani *et al.* 2003[89]	Wimo *et al.* 2003[56]	AD2000 Collaborative Group 2004[43]
8. Costs and consequences valued credibly	*Costs:* ✓ Japanese government fees and prices used for medical and long-term care. Drug costs based on Japanese drug price *Consequences:* ?/× Canadian QoL instrument may not reflect the sense of values in Japan	*Costs:* ?/✓ From the Polish Alzheimer's Society *Consequences:* ? Calculated from model in terms of years in a non-severe AD state over 5 years, with some estimates from unknown sources	*Costs:* ✓ French cost sources used. Unpaid care was considered a substitute to paid care and valued conservatively *Consequences:* ? Calculated from model in terms of percentage population in different care settings over 3 years	*Costs:* ?/✓ Unit prices obtained from Swedish national statistics and scientific publications. All costs reported in 1999 values *Consequences:* ✓ No modelling required as no extension beyond the 1-year trial period	*Costs:* ?/✓ Unit prices obtained from UK national average unit costs. Costs believed to be 2000 values although this not stated explicitly *Consequences:* ✓ No modelling required as no extension beyond the trial period
9. Differential timing considered	✓ 3% discount rate for costs and QALYs	? Not reported	× Undiscounted	× Undiscounted (1-year only)	× Undiscounted
10. Incremental analysis performed	✓	✓	?	×	×
11. Sensitivity analysis performed	✓ One-way	✓ One-way	✓ One-way	✓ One-way	✓ Multivariate
12. Modelling conducted reasonably	Markov model used	Markov model used	Spreadsheet model used	Simple cost consequence model	

? means unclear or unknown
✓ means item included or judged as acceptable to be internally valid
× means factor not included or judged as unacceptable to be internally valid.

External validity of donepezil economic studies

Item/study	Stein 1997[81]	Stewart *et al.* 1998[82]	Jönsson *et al.* 1999[84]	O'Brien *et al.* 1999[85]	Neumann *et al.* 1999[86]	Ikeda *et al.* 2002[87]	Sobolewski *et al.* 2002[88] (Abstract)	Fagnani *et al.* 2003[89]	Wimo *et al.* 2003[56]	AD 2000 2004[43]
1. Patient group – are the patients in the study similar to those of interest in England and Wales?	✓/?	✓/?	?	?	?	?	?	?	?	✓/?
2. Health care system/ setting – comparability to England and Wales?; comparability of available alternatives?; similar levels of resources?; institutional arrangements comparable?	✓/?	✓/?	×	×	×	×	?	?	×	✓
3. Treatment – comparability with clinical management?	✓/?	✓/?	?	?	?	?	?	?	?	✓
4. Resource costs – comparability between study and setting/ population of interest?	✓/?	✓/?	×	×	×	×	×	×	×	✓/?

Note ? means unclear or unknown
✓ means judged item suitable to generalise to England and Wales with or without some re-adjustment.
× means factor judged not suitable as either not possible to see how an adjustment could be made easily in short/medium term or relevant data unavailable.

Economic evaluations of rivastigmine

Internal validity of rivastigmine studies

Note ? means unclear or unknown.

✓ means item included or judged as acceptable to be internally valid.

× means factor not included or judged as unacceptable to be internally valid.

Item/study	Stein 1998[90]	Fenn and Gray 1999[91]	Hauber *et al.* 2000[92]	Hauber *et al.* 2000[93]	Brooks and Deal 2000[94]
1. Well defined question	✓	✓	✓	✓	✓
2. Clear description alternatives	× Rivastigmine dose unclear (?6–12 mg), versus placebo	✓ Rivastigmine low dose (1–4 mg/day), and high dose (6–12 mg/day), versus placebo	?/✓ Rivastigmine (unspecified dose) versus placebo	✓ Rivastigmine pooled low (1–4 mg/day) and high (6–12 mg/day) doses, versus no treatment	× Rivastigmine dose not stated and no description of alternative
3. Reasonable study type	CUA – by simple calculation	CEA	CEA	CEA	CEA
		These studies all use the parametric hazard model of disease progression developed by Fenn and Gray[91]			
4. Effectiveness established	✓ Cites three phase II trials and 4 phase III trials (only 1, Corey-Bloom *et al.*[57] published in full). Presents summary findings.	?/✓ Cites Rösler[58] and Corey-Bloom.[57] Presents a short narrative summary of the findings stating clear evidence of efficacy provided. No numerical results presented	× Cites Rösler[58] and Corey-Bloom[57] plus Fenn and Gray[91] but does not present any findings from the trials	× Cites Rösler[58] and Corey-Bloom[57] plus Fenn and Gray[91] but does not present any findings from the trials	× Cites two phase III clinical trials but does not provide reference or any findings from the trials

continued

Item/study	Stein 1998[90]	Fenn and Gray 1999[91]	Hauber *et al.* 2000[92]	Hauber *et al.* 2000[93]	Brooks and Deal 2000[94]
5. Effectiveness estimates related to population risks	× Trial populations were located in Europe (various countries including UK), Canada, USA, South Africa and Australia. No discussion of how the trial patient characteristics relate to the UK population	? Trial population data from Rösler[58] and Corey-Bloom[57] in Europe (Austria, France, Germany and Switzerland) and North America (Canada and USA) used to estimate disease progression and duration of health benefits due to treatment. Data from a UK survey of AD patients in long-term care[126] and trial data used to estimate the probability of institutionalisation for 3 severity levels of AD. No discussion of how the trial patient characteristics relate to the UK population	?/✓ Trial population data from Rösler[58] and Corey-Bloom[57] in Europe (Austria, France, Germany and Switzerland) and North America (Canada and USA) used to estimate disease progression and duration of health benefits due to treatment. Data from a USA AD registry (CERAD) used to estimate the probability of institutionalisation for 3 severity levels of AD. Authors point out that cost estimates were derived from a small sample of AD patients in California[197] so the results might not be representative of a larger more diverse patient population	?/✓ Trial population data from Rösler[58] and Corey-Bloom[57] in Europe (Austria, France, Germany and Switzerland) and North America (Canada and USA) used to estimate disease progression and duration of health benefits due to treatment. Probability estimates of institutionalisation for 4 severity levels of AD came from the Canadian study of Health and Aging (CSHA)[195]	? No reported information

continued

Item/study	Stein 1998[90]	Fenn and Gray 1999[91]	Hauber *et al.* 2000[92]	Hauber *et al.* 2000[93]	Brooks and Deal 2000[94]
6. Relevant costs and consequences identified	*Costs:* ✓/? Drug costs used in the analysis. Direct costs for outpatient and memory clinic visits, and CT scanning estimated but not used in the analysis. Some formal care costs (residential care, social services) and all informal care costs were not included *Consequences:* ? Estimates QALY gains (method unclear). NNTs for clinically important outcomes summarised. Carer quality of life not included	*Costs:* ✓/? Cites paper reporting estimates of unit costs of home and institutional care in the UK.[126] Home care costs included GP consultations, outpatient visits, day care, respite care, home care, meals on wheels and short-term hospitalisations. Informal care and drug treatment costs not included *Consequences:* ? Model estimates cost savings as a consequence of delayed disease progression	*Costs:* ? Cites the paper reporting a cost estimate for institutional care and an equation for calculating cost of community care according to MMSE score.[197] The costs contributing to these estimates are not described. Drug treatment costs are not included *Consequences:* ? Model estimates cost savings as a consequence of delayed disease progression	*Costs:* ✓/? Cites paper reporting costs estimated from the CSHA according to disease severity.[195] Costs included nursing home care, medications, caregivers' use of community support services, and unpaid caregiver time. Drug treatment costs also included *Consequences:* ? Model estimates cost savings as a consequence of delayed disease progression. QALY threshold analysis presented	*Costs:* ?/✗ Direct and indirect costs included but no details provided *Consequences:* ? Decrease in likelihood of institutionalisation
7. Costs and consequences measured accurately	*Costs:* ✓/? Use of clinic visits and CT scanning services estimated. No information given to support the estimates of service use *Consequences:* ? QALYs calculated using three assumptions: 1) cognitive decline measures disease progression, 2) mortality unaffected by rivastigmine treatment, and 3) after 6 months rivastigmine has no effect on rate of disease progression. Possible errors in the measurement of consequences	*Costs:* ? Average costs for home care and institutional care in the UK were from a previous study of AD costs[126] *Consequences:* ? MMSE is used as a measure of disease progression. To obtain estimates of when patients moved from one MMSE level to the next (so that time to disease progression could be calculated for use in the model) linear interpolation between trial visit dates was used. Possible errors in the measurement and calculation of consequences	*Costs:* ? Average costs for home care and institutional care were from a previous study of AD costs in the USA (cited)[197] *Consequences:* ? The MMSE score of patients at the point of institutionalisation was estimated by adjusting the observed MMSE scores of institutionalised patients using the average decline in MMSE score of patients with AD in CERAD data. The likelihood that an individual would be institutionalised for a given MMSE score was then modelled. Possible errors in the measurement and calculation of consequences	*Costs:* ?/✓ Estimates of the costs of caring for people with different severity levels of AD were from a previous study (cited).[198] Costs were adjusted to account for the oversampling in the CSHA data of persons in institutions *Consequences:* ? Detailed description of the model not provided. Cites Fenn and Gray,[91] and Hauber[92]	*Costs:* ?/✗ No reported information *Consequences:* ?/✗ No reported information

continued

Item/study	Stein 1998[90]	Fenn and Gray 1999[91]	Hauber *et al.* 2000[92]	Hauber *et al.* 2000[93]	Brooks and Deal 2000[94]
8. Costs and consequences valued credibly	*Costs:* ✓ UK cost sources used. Source of drug costs not provided, unclear which year these relate to. Costs of outpatient and memory clinic visits, and CT scan calculated as the average of ECR prices in 1996/7 (this data not included in final analysis) *Consequences:* × QALYs gained from the use of rivastigmine estimated using the IHQL (not validated for valuing cognitive impairment). No rationale given for the IHQL values used	*Costs:* ? Sources of costs used in paper cited, not summarised by authors. Cost savings estimated by the model *Consequences:* ? Days saved from a more severe disease stage are predicted from the model and illustrated by the area between two curves in a figure. No numerical values are reported	*Costs:* ? Sources of costs used in the paper cited not summarised by authors. Cost savings are estimated by the model *Consequences:* ? Delay in cognitive decline is predicted by the model for 6-month, 1-year and 2-year time horizons	*Costs:* ?/✓ Costs estimated in a cited paper used CSHA data on utilisation of resources. Average daily cost savings are estimated by the model *Consequences:* ? Delay in cognitive decline is predicted from the model, over 6-month, 1-year and 2-year time horizons. QALY thresholds are reported based on 2 arbitrary monetary values for a gain of 1 QALY	Costs: ?/× No reported information Consequences: ?/× No reported information
9. Differential timing considered	✓ 6% discount rate for costs	× Undiscounted	✓ 3% discount rate applied to second year cost savings	✓ 3% discount rate applied to all second year costs and cost savings	? Not reported
10. Incremental analysis performed	✓	✓	✓	✓	? Not reported
11. Sensitivity analysis performed	✓ Two-way	×	×	×	? Not reported
12. Modelling conducted reasonably	Simple decision analysis model	Survival analysis model	Survival analysis model	Survival analysis model	? Insufficient detail

Note ? means unclear or unknown
✓ means item included or judged as acceptable to be internally valid.
× means factor not included or judged as unacceptable to be internally valid.

External validity of rivastigmine economic studies

Item/study	Stein 1998[90]	Fenn and Gray 1999[91]	Hauber et *al.* 2000[92]	Hauber et *al.* 2000[93]	Brooks and Deal 2000[94] (abstract)
1. Patient group – are the patients in the study similar to those of interest in England and Wales?	?/✓	?/✓	?	?	?
2. Health care system/setting – comparability to England and Wales?; comparability of available alternatives?; similar levels of resources?; institutional arrangements comparable?	✓/?	✓/?	×	×	×
3. Treatment – comparability with clinical management?	?	?	?	?	?
4. Resource costs – comparability between study and setting/population of interest?	✓/?	✓/?	×	×	?

Note ? means unclear or unknown
✓ means judged item suitable to generalise to England and Wales with or without some re-adjustment.
× means factor judged not suitable as either not possible to see how an adjustment could be made easily in short/medium term or relevant data unavailable.

Economic evaluations of galantamine

Internal validity of galantamine studies

Note ? means unclear or unknown.

✓ means item included or judged as acceptable to be internally valid.

× means factor not included or judged as unacceptable to be internally valid.

Item/study	Getsios *et al.* (2001)[95]	Garfield *et al.* (2002)[96]	Caro *et al.* (2002)[97]	Migliaccio-Walle *et al.* (2003)[98]	Ward *et al.* (2003)[99]
1. Well defined question	✓	✓	✓	✓	✓
2. Clear description alternatives	✓ Galantamine 24 mg, versus non-pharmacological treatment	✓ Galantamine 12 mg, three times daily (36 mg) versus non-pharmacological treatment	× Not clearly stated. Study refers to two clinical trials, but galantamine dose not stated in the paper. Comparator is placebo/no drug treatment	✓ Galantamine 16 mg, and 24 mg, versus non-pharmacological treatment	✓ Galantamine 16 mg, and 24 mg, versus non-pharmacological treatment
3. Reasonable study type	✓ CEA – using a CE model	✓ CEA – using a CE model	✓ CEA – using a CE model	✓ CEA – using a CE model	✓ CEA – using a CE model, presenting CUA results
4. Effectiveness established	✓	×	×	×	×
All studies (1–5) cite trials reported by Raskind *et al.*,[61] Wilcock *et al.*,[64] and Tariot *et al.*,[63] to establish effectiveness Getsios *et al.* present summary findings from the trials, but all other studies failed to present findings from the clinical trials cited.					

continued

Item/study	Getsios *et al.* (2001)	Garfield *et al.* (2002)	Caro *et al.* (2002)	Migliaccio-Walle *et al.* (2003)	Ward *et al.* (2003)
5. Effectiveness estimates related to population risks	? Estimates of effectiveness not directly related to a Canadian population. The model used applies trial population data and characteristics from Raskind et al. (USA patients), and Wilcock *et al.* (multinational trial, 8 countries including Canada), as a baseline group. The risk associated with the economic endpoint of requiring FTC are modelled based on the experiences of a cohort of US patients	× Estimates of effectiveness not related to a UK population. The model used applies trial population data and characteristics from Raskind *et al.* (USA patients), Wilcock *et al.* (multinational trial, 8 countries including Sweden), as a baseline group. No discussion of how the trial patient characteristics relate to Sweden. The risk associated with the economic endpoint of requiring FTC are modelled based on the experiences of a cohort of US patients	× Estimates of effectiveness not related to a UK population. The model used applies trial population data and characteristics from Raskind *et al.* (USA patients), and Wilcock *et al.* (multinational trial, 8 countries including Netherlands), as a baseline group. No discussion of how the trial patient characteristics relate to Netherlands. The risk associated with the economic endpoint of requiring FTC are modelled based on the experiences of a cohort of US patients	? The model used applies trial population data and characteristics from two USA trials as a baseline group; Raskind *et al.* (USA patients), and Tariot *et al.* (USA patients). These trials reflect a USA patient group, but the trial patients (with specific inclusion/exclusion criteria) may not reflect the general 'in-practice' treatment/patient group. The risk associated with the economic endpoint of requiring FTC are modelled based on the experiences of a cohort of US patients	× Estimates of effectiveness not related to a UK population. The model used applies trial population data and characteristics from Raskind *et al.* (USA patients), Wilcock *et al.* (multinational trial, 8 countries including UK) and Tariot et al (USA patients), as a baseline group. No discussion of how the trial patient characteristics relate to UK. The risk associated with the economic endpoint of requiring FTC are modelled based on the experiences of a cohort of US patients
6. Relevant costs and consequences identified	*Costs:* ✓ Formal care costs, including medical care costs and paid home help. Informal care costs not included. *Consequences:* ? Estimate patient benefits of time until patients require FTC, and QALY gains	*Costs:* ✓ Direct costs covered by health insurance included, including acute care, health professional visits, nursing home care and other paid caregivers. *Consequences:* ? Estimate patient benefits of time until patients require FTC	*Costs:* ✓ Formal care costs, including medical care costs and paid home help (including drug costs, nursing home care, outpatient care, general hospital care and psychiatric hospital care). Informal care costs not included. *Consequences:* ? Estimate patient benefits of time until patients require FTC, time to death, and QALY gains	*Costs:* ✓ Costs included all paid formal care costs (incl. nursing home, hospitalisation, physician visits, nursing care, emergency dept. visits, galantamine costs); excluded costs borne by patients or their families/caregivers *Consequences:* ? Estimate patient benefits of time until patients require FTC	*Costs:* ✓ Direct costs including medical and social service costs, associated with model states, are used in the analysis. *Consequences:* ? Estimate patient benefits of time until patients require FTC, time to death, and QALY gains

continued

Item/study	Getsios et *al*. (2001)	Garfield et *al*. (2002)	Caro et *al*. (2002)	Migliaccio-Walle et *al*. (2003)	Ward et *al*. (2003)
7. Costs and consequences measured accurately	*Costs:* ? First 6-month costs based on clinical trial data (Wilcock *et al*.), using a resource profile for 114 Canadian trial patients (for mild–moderate AD), and all trial patients for moderate AD. Resource use data for the cost estimates were from various sources. A community follow-up survey, using GDS scores to calculate resource use by health state i.e. full-time community care (FTC) and pre-FTC, was used for community estimates on FTC and pre-FTC. Estimated from a published survey that 73.5% of patients requiring FTC would reside in nursing homes. Limited detail offered to support cited sources *Consequences:* ? ADAS-cog scores from trial data transformed to MMSE scores, then to mMMSE scores, and used to model the endpoint of time to FTC. Possible measurement errors in the measurement and calculation of consequences	*Costs:* ?/× Estimated from a published survey that 85% of patients requiring FTC would reside in an institution. Patients in nursing homes estimated at 68% of patients requiring FTC, and 81% of patients living in institutions. Estimates for resource use by FTC outside of an institution derived from patients in 'group living' facility. Limited detail offered to support estimates and cited sources *Consequences:* ? ADAS-cog scores from trial data transformed to MMSE scores, then to mMMSE scores, and used to model the endpoint of time to FTC. Possible measurement errors in the measurement and calculation of consequences	*Costs:* ?/× First 6-month costs based on clinical trial data. It is unclear how average pre-FTC costs are calculated. Estimates for proportions of patients in institutions not justified in the paper; the authors cite three studies but do not discuss. The cost for FTC in the community is estimated at 10.73% of the cost of institutional care, based on a cited study, but no detail is offered to support the citation *Consequences:* ? ADAS-cog scores from trial data transformed to MMSE scores, then to mMMSE scores, and used to model the endpoint of time to FTC. Possible measurement errors in the measurement and calculation of consequences	*Costs:* ?/× Cost estimates calculated for components of the model (initial 6-months, pre-FTC, FTC comm., FTC nursing home). First 6-month costs based on clinical trial data. Assumptions made over proportions of patients entering a nursing home. Limited information provided on the resource use profiles used for cost estimates. Resource use for pre-FTC from trial data. Resource use for FTC from published US study. Authors assumed 41% of FTC patients would be in the community, 59% in nursing home *Consequences:* ? ADAS-cog scores from trial data transformed to MMSE scores, then to mMMSE scores, and used to model the endpoint of time to FTC. Possible measurement errors in the measurement and calculation of consequences	*Costs:* ? Community and institutional care costs estimated from UK surveys in cognitively impaired patients. These surveys were dated 1985 and 1988, but judged by study authors to be relevant for data on the proportion of patients in a residential care setting. Data from 2000 and 2001 published sources used to determine percentage of patients in different residential settings *Consequences:* ? ADAS-cog scores from trial data transformed to MMSE scores, then to mMMSE scores, and used to model the endpoint of time to FTC. Possible measurement errors in the measurement and calculation of consequences

continued

Item/study	Getsios et *al*. (2001)	Garfield et *al*. (2002)	Caro et *al*. (2002)	Migliaccio-Walle et *al*. (2003)	Ward et *al*. (2003)
8. Costs and consequences valued credibly	*Cost:* ?/× Cost data and assumptions are reported by the authors, but support (citations) for unit costs not reported. Cost per hospitalisation day used is $1012.38; this appears a high estimate *Consequences:* ? Consequences are calculated from the economic model in terms of time delay/period. QALYs used in analysis, with data from a published study (pre-FTC at 0.60, FTC at 0.34)	*Cost:* ?/× Authors state all costs are based on estimates from literature (published?), but specific data are not reported. Authors detail the monthly cost estimated derived for health states used in the model (pre-FTC and FTC). Swedish cost-of-illness study (for 1991) used to estimate costs for FTC *Consequences:* ? Consequences are calculated from the economic model in terms of time delay/period	*Cost:* ? The study states individual unit cost data for many items, but does not specifically support the data with a source. The study states that unit cost data were obtained from different sources, providing 7 references (general unpublished health service or industry sources) *Consequences:* ? Consequences are calculated from the economic model in terms of time delay/period. QALYs used in analysis, with data from a published study (pre-FTC at 0.60, FTC at 0.34)	*Cost:* ?/× Various cost data sources used. Initial 6-month data from inpatient admissions database, Medicare fee schedules, telephone surveys and other published sources (but no specific item costs reported). Pre-FTC cost data sources were the same as the initial 6-months. For FTC cost databases info cited, but no detail provided on cost data items used in estimates *Consequences:* ? Consequences are calculated from the economic model in terms of time delay/period	*Cost:* ✓ UK cost sources used *Consequences:* ? Health state utilities for states of pre-FTC and FTC derived from a published study (pre-FTC at 0.60, FTC at 0.34)
9. Differential timing considered	✓ 3% discount rate for costs and consequences	✓ ? 3% discount rate for costs discounting/discount rate for consequences not stated	✓ 5% for costs and QALYs	✓ 3% discount rate for costs and consequences	✓ 6% discount rate for costs 1.5% discount rate for consequences
10. Incremental analysis performed	✓	✓	✓	✓	✓
11. Sensitivity analysis performed	✓ One-way sensitivity analysis reported	✓ One-way sensitivity analysis reported	✓ One-way sensitivity analysis reported	✓ One-way sensitivity analysis reported	✓ One-way sensitivity analysis reported
12. Modelling conducted reasonably	AHEAD model used†	AHEAD model used†	AHEAD model used†	AHEAD model used†	AHEAD model used†

Note ? means unclear or unknown.
✓ means item included or judged as acceptable to be internally valid.
× means factor not included or judged as unacceptable to be internally valid.

Galantamine

Item/study	Getsios et *al.* (2001)[95]	Garfield et *al.* (2002)[96]	Caro et *al.* (2002)[97]	Migliaccio-Walle et *al.* (2003)[98]	Ward et *al.* (2003)[99]
1. Patient group – are the patients in the study similar to those of interest in England and Wales?	?	?	?	?	✓/?
2. Health care system/setting – comparability to England and Wales?; comparability of available alternatives?; similar levels of resources?; institutional arrangements comparable?	×	?	×	×	✓/?
3. Treatment – comparability with clinical management?	?	?	?	?	✓/?
4. Resource costs – comparability between study and setting/population of interest?	×	×	×	×	✓

Note ? means unclear or unknown
✓ means judged item suitable to generalise to England and Wales with or without some re-adjustment.
× means factor judged not suitable as either not possible to see how an adjustment could be made easily in short/medium term or relevant data unavailable.

Economic evaluations of memantine

Internal validity of memantine studies

Item/study	François et *al.*[103]	Jones et *al.*[104]
1. Well defined question	✓	✓
2. Clear description alternatives	✓ Memantine versus no pharmacological treatment (placebo)	✓ Memantine versus no pharmacological treatment (placebo)
3. Reasonable study type	✓ CEA – using a CE model	✓ CEA/CUA – using a CE model
4. Effectiveness established	× Cites trials by Reisberg *et al.*, and Ferris *et al.*, but does not present outline findings from these studies	× Cites trial by Reisberg *et al.*, but does not present outline findings from the study
5. Effectiveness estimates related to population risks	× Effectiveness estimates are not related to population risks. No discussion of how the trial patient characteristics relate to Finland. Study uses transition probabilities from trial data, and other sources, to model population risks, and does not directly use the primary effectiveness variables from clinical trials in analysis	× Effectiveness estimates are not related to population risks. No discussion of how the trial patient characteristics relate to UK. Study uses transition probabilities from trial data, and other sources, to model population risks, and does not directly use the primary effectiveness variables from the clinical trial in analysis
6. Relevant costs and consequences identified	*Costs:* ✓/? Societal perspective adopted, including costs for informal care. This may be appropriate for the Finnish analysis, but not appropriate for the NHS and PSS perspective of the NHS/NICE. Relevant NHS and PSS cost items identified, but included as part of an overall cost estimate, comprising other societal costs *Consequences:* ? Estimate patient benefits of time in independent state, time to institutionalisation. No final health outcomes e.g. QALYs used. Some analysis related to cost per year of independency gained	*Costs:* ? Costs per cycle by dependency and location were used in the model. Where cost components are discussed they appear to be consistent with an NHS and PSS perspective; however, cost estimates per cycle appear high and may include cost components that do not fall on the NHS and PSS budget *Consequences*: ? Estimate patient benefits of time in independent state, time to institutionalisation and QALY gains

continued

Item/study	François et *al.*[103]	Jones et *al.*[104]
7. Costs and consequences measured accurately	*Costs:* ✓/? Resource use information was from a Finnish epidemiological study	*Costs:* ?/× Resource use from a cited study (LASER-AD Study), but the generalisability of the data from the cited source is in question. Insufficient detail presented in the manuscript
	Consequences: ? Consequences estimated from the disease progression model developed as part of this study	*Consequences:* ?/× Insufficient detail presented in the manuscript
8. Costs and consequences valued credibly	Costs:✓/? Unit cost data from published Finnish source/s	*Costs:* ?/× Cited source for unit cost data from credible source, but estimate of resource use may not be generalisable. Insufficient detail presented in the manuscript. Cost data for institutional care not specified and estimates reported for institutional care seem higher than other published estimates
	Consequences: ? Consequences estimated from the disease progression model developed as part of this study. No monetary or utility value used	*Consequences:* ?/× Insufficient detail presented in the manuscript
9. Differential timing considered	✓ 3.5% discount rate for costs and consequences	✓ 3.5% discount rate for costs and consequences
10. Incremental analysis performed	✓	✓
11. Sensitivity analysis performed	✓ Probabilistic sensitivity analysis reported, and other uncertainties via one-way sensitivity analysis	✓ Probabilistic sensitivity analysis undertaken, but not reported in any detail. Subgroup analysis reported in outline
12. Modelling conducted reasonably	? Markov-type model, using transition probabilities for severity, dependency and location (multiplicative probabilities). The link between effectiveness and disease progression is an indirect link	? Markov-type model, using transition probabilities for severity, dependency and location (multiplicative probabilities). The link between effectiveness and disease progression is an indirect link

Note ? means unclear or unknown.
✓ means item included or judged as acceptable to be internally valid.
× means factor not included or judged as unacceptable to be internally valid.

External validity of memantine economic studies

Item/study	François et *al.*[103]	Jones et *al.*[104]
1. Patient group – are the patients in the study similar to those of interest in England and Wales?	?/✓	?/✓
2. Health care system/setting – comparability to England and Wales?; comparability of available alternatives?; similar levels of resources?; institutional arrangements comparable?	✓/?	?
3. Treatment – comparability with clinical management?	?	?
4. Resource costs – comparability between study and setting/population of interest?	✓/?	?

Note ? means unclear or unknown.
✓ means judged item suitable to generalise to England and Wales with or without some re-adjustment.
× means factor judged not suitable as either not possible to see how an adjustment could be made easily in short/medium term or relevant data unavailable.

Appendix 14

Summary results from cost-effectiveness literature

Summary reporting of cost-effectiveness in included studies: by product

Summary results: economic evaluations for donepezil

Stein 1997[81]
Results presented as costs per QALY based on drug costs only. A gain of between 0.05 and 0.08 QALYs is estimated. For a benefit of 0.05 QALYs the cost per QALY of a 5 mg/day dose for 2 years is predicted to be £34,640, rising to £79,560 for 5 years, £117,280 for 8 years and to £139,020 for 10 years. Increasing the benefit to 0.08 QALYs reduces the cost per QALY to £21,383 over 2 years, rising to £85,815 for 10 years of treatment.

Sensitivity analysis:
Donepezil 10 mg/day: the evidence for the marginal benefit is less clear than that for the disbenefits associated with the higher dose. With a benefit of 0.05 QALYs the cost per QALY of a 10 mg/day dose for 2 years is £48,500, rising to £194,720 for 10 years. Assuming a higher QALY gain of 0.08 reduces the cost per QALY to £29,938 for 2 years, and £120,198 for 10 years.

Stewart *et al.*, 1998[82]
Results presented as incremental investment in treatment required to achieve an extra year spent in a non-severe AD state. Treatment groups described as being almost cost neutral because over the 5-year time horizon costs are raised only very slightly.

Mild AD:
Donepezil 10 mg: an incremental cost (over placebo) of £5,698 is required to obtain an extra year in a non-severe AD state; the incremental cost vs the 5 mg dose is £4,451.
Donepezil 5 mg: an incremental cost (over placebo) of £7,048 is required to obtain an extra year in a non-severe AD state.

Moderate AD:
Donepezil 10 mg: an incremental cost (over placebo) of £3,562 is required to obtain an extra year in a non-severe AD state; there is no incremental benefit versus the 5 mg donepezil dose.
Donepezil 5 mg: an incremental cost (over placebo) of £1,210 is required to obtain an extra year in a non-severe AD state.

Sensitivity analysis:
Variations in the discount rate: as the discount rate is altered the relative positions of the subgroups do not alter.
Mortality rate lowered to 30% over 3 years: as the mortality rate falls, more patients remain alive at later stages and continue to incur costs. However, the incremental costs (over placebo) for an additional year in a non-severe AD state reduce slightly to £4,955 and £5,328 for 10 mg and 5 mg donepezil groups respectively in mild AD. In moderate AD the incremental costs for the 10 mg and 5 mg doses are £3,372 and £942 respectively.

Jönsson *et al.*, 1999[84]
Results presented as a cost-effectiveness comparison based on patient years in non-severe AD states, and cumulative 5-year costs of care. Treatment with donepezil 5 mg and 10 mg produced cost savings. The lower dose was reported as the most cost-effective option because the effectiveness of the two doses was not statistically different in the clinical trial.

Kungsholmen analysis:
Donepezil 5 mg associated with expected cost savings over 5 years (undiscounted) over no treatment of SEK 15,561 and a gain of 0.5222 years in non-severe AD states. Donepezil 10 mg associated with expected cost savings of SEK 3,426 and a gain of 0.5424 years in non-severe AD states.

Within trial analysis:
Donepezil 5 mg associated with expected cost savings over 5 years (undiscounted) over no treatment of SEK 277,631 and a gain of 0.7175 years in non-severe AD states. Donepezil 10 mg associated with expected cost savings of SEK 237,031 and a gain of 0.8484 years in non-severe AD states.

Sensitivity analysis:
Patients switched from with-treatment to no-treatment transition probabilities after a number of cycles: donepezil treatment was cost saving even when the no-treatment transition probabilities were introduced after the first cycle of the model; the impact on time spent in non-severe AD states is not reported.

continued

O'Brien *et al.*, 1999[85]
Results presented as expected cost per patient and the expected years per patient in a non-severe AD state over 5 years. Incremental differences presented for the combined AD group only. Reports that the best estimate is that donepezil will both save money and improve patient outcomes.

Mild AD:
Donepezil 5 mg associated with an increase in societal costs of Can $2100 over no treatment and an expected gain of 0.1 years per patient in a non-severe AD state.

Mild to moderate AD:
Donepezil 5 mg associated with a cost saving of Can $1497 and an expected gain of 0.17 years per patient in a non-severe AD state.

Moderate AD:
Donepezil 5 mg associated with a cost saving of Can $5589 and an expected gain of 0.44 years per patient in a non-severe AD state.

Combined groups:
Societal costs of Can $80,305 with 5 mg donepezil treatment are less than the no treatment costs of Can $81,187, representing a saving of Can $882, while the expected years per patient in a non-severe AD state with donepezil treatment are 2.41 in comparison to 2.21 with no treatment, giving an incremental benefit of 0.2 years in non-severe AD states.

Sensitivity analysis:
Assessed for combined groups only.

Mortality rate:
With 100% survival at 5 years cost savings increase to Can $1244 and years in non-severe AD states increase to 0.24. As survival falls to 0%, cost savings and years in non-severe AD states decrease to Can $268 and 0.12 years.

Discount rate:
Decreasing the discount rate increases cost savings, and benefits either increase or are maintained. Raising the discount rate to 7% decreases cost savings to Can $820 and years in a non-severe AD state to 0.19 years.

Treating severe AD with donepezil:
Cost savings replaced by an incremental cost of Can $1554 although 0.18 years in a non-severe AD state were still achieved.

Monthly rather than quarterly dispensing charges:
Decreased cost savings to Can $759, no effect on benefits.

Alternative MMSE aggregation weights from CSHA:
Increased both cost savings and years in non-severe AD states (Can $1,447, 0.22 years).

Increasing the value of caregiver time:
Decreased cost savings to Can $826, no effect on benefits.

Neumann *et al.*, 1999[86]
Results presented as cost per QALY over 18 months. Reports that the cost of donepezil may be partially offset by a reduction in the costs of care.

Mild AD in the community:
CE ratio of donepezil treatment is $9300/QALY for the 18 month base-case time horizon. Drug treatment only starts to become cost saving if drug effect is assumed to persist for 24 months. Assuming drug treatment effects persist for only 6 or 12 months results in costs per QALY of $160,000 or $32,000 respectively.

Moderate AD in the community:
CE ratio of donepezil treatment is $76,000/QALY (base-case). The model does not predict cost savings regardless of the duration of drug effect. The CE ratio is predicted to fall below $50,000/QALY when the drug effect is assumed to persist beyond 24 months. Assuming drug treatment effects persist for only 6 or 12 months leads to costs per QALY of $440,000 or $140,000 respectively.

Sensitivity analysis:
Presented for the assumption of 18 month duration of drug effect and reported to be similar for the 6- and 12-month time horizons. Cost savings are only reported twice, for the mild AD/community group in both cases; when the risk ratio estimated in the proportional hazards model is one standard error better than the mean the drug is estimated to save $319 and 0.038 QALYs, and when the total treatment drug cost is $1,000 cost savings also occur (level of savings not reported). CE ratios are generally highly sensitive to assumptions about drug duration, slightly sensitive to varying rates of continuation, and insensitive to the following: varying the absolute magnitude of preference weights or the inclusion of caregiver's preference weights; changes in the discount rate; the rate of disease progression; and the rate of nursing home placement.

continued

Ikeda *et al.*, 2002[87]
Results presented as ICERs (cost per QALY). Reports that donepezil is dominant with cost savings and health outcomes.

Mild AD:
Incremental CE ratio of donepezil treatment is cost saving (18-month base-case time horizon). Donepezil associated with costs of 2,596,668 yen and a gain of 0.579 QALYs in comparison to costs of 2,600,541 yen and 0.506 QALYs gained in the conventional therapy group.

Moderate AD:
Incremental CE ratio of donepezil treatment is cost saving (18-month base-case time horizon). Donepezil associated with costs of 4,617,426 yen and a gain of 0.310 QALYs versus 4,882,970 yen and 0.227 QALYs in the conventional therapy group.

Sensitivity analysis:
Presented for assumptions of drug efficacy from 6 months to 2 years.

Mild AD:
Donepezil is not cost saving when drug efficacy is assumed to last only 1 year or 6 months; the incremental cost/QALY becomes 1,462,658 yen and 6,756,958 yen respectively.

Moderate AD:
Donepezil is cost saving except when drug efficacy is assumed to last only 6 months. Then the incremental cost/QALY is 305,476 yen.

Sobolewski *et al.*, 2002[88]
Results presented as expected costs, years in a non-severe AD state, and incremental costs per additional years in a non-severe AD state over 5 years. Reports that donepezil is most cost-effective in mild AD and cost-effectiveness does not increase if treatment lasts longer than 6 months.
Incremental cost of additional year in a non-severe AD state per treated person is 13,260 zl/year

Sensitivity analysis:
Percentage of mild AD patients in the model raised from 50% to 75% or 100%: incremental cost-effectiveness of additional year in a non-severe AD state becomes 9457 zl/year and 6776 zl/year respectively.

Fagnani *et al.*, 2003[89]
Results presented as place of care and net costs of care over 3 years. Reports that donepezil is a cost-effective option in the management of mild to moderate AD.

Place of care:
Model predicts home care for 74% of surviving treated patients compared with 68% of untreated patients. Of the remaining patients 12% of treated patients and 15% of untreated patients would be in non-medical institutions and 13% of treated and 18% of untreated patients would be in medical institutions. Treated patients spent more time at home (30 months per patient) compared with untreated patients (28 months per patient).
Total net costs of care over 3 years: Eur 42,720 for treated patients, and Eur 53,206 for untreated patients.

Cost savings:
Approximately Eur 3500 per year with donepezil, mainly due to savings in unpaid caregiver time.

Sensitivity analysis:
Unit costs and estimates of time spent by caregivers and professionals, used for paid and unpaid assistance, were the inputs that the 3-year costs calculated by the model were most sensitive to. Results are much less sensitive to variations in hostel cost in residential care of various categories or to costs of general medical expenditures.

Wimo *et al.*, 2003[56]
Results presented as cost savings per patient based entirely on a 1-year clinical trial (no modelling required) with reduced deterioration in basic and instrumental activities of daily living. Reports that cognitive and functional benefits of donepezil treatment are realised with no increase in costs to society compared with placebo treatment over 1 year.
Overall cost saving with donepezil of SEK 9190 (US$1097) made up of patient direct costs (a cost in comparison to placebo of SEK 2438,), caregiver direct costs (a saving vs placebo of SEK 2973), and caregiver time costs (a saving vs placebo of SEK 8655).

Sensitivity analysis:
The impact of the following costs was assessed: costs imputed for patients withdrawing from the study; non-AD hospitalisation costs; patient accommodation costs; and caregiver time costs. The magnitude of the treatment differences was altered but not the direction of the results.

continued

AD2000 2004[43]
Results presented as estimated annual costs per patient based on resource use during the clinical trial (no modelling required) with no significant benefits in institutionalisation or progression of disability (the two primary effectiveness outcomes). Reports that donepezil is not cost effective with benefits below minimally relevant thresholds.
Overall cost with donepezil of £498 which does not include the cost of donepezil or institutionalisation.

Sensitivity analysis:
Multivariate analyses indicated that improvements in functional ability with donepezil (assuming the maximum benefit compatible with the trial data) would not delay institutionalisation sufficiently to offset the costs of the drug. The authors state that the impact of variations in key variables was tested but this is not reported in detail. There were no plausible valuations of unpaid caregiver time that would offset the higher costs of formal care seen with donepezil.

Summary results: economic evaluations for rivastigmine

Stein 1998[90]
Results all presented as cost per QALY based on drug costs only. The analysis assumes no survival advantage. Costs but not benefits were discounted.
The benefits of rivastigmine were estimated to be between 0.05 and 0.08 QALYs. For a benefit of 0.05 QALYs the cost per QALY for one year is predicted to be £16,420 rising to £31,900 for 2 years and to £73,320 for five years. Increasing the benefit to 0.08 QALYs reduces the cost per QALY to £10,263 over one year, rising to £45,825 for five years of treatment.

Sensitivity analysis:
When estimates for non-drug treatment were included, QALY estimates ranged from £14,543 to £88,915.

Fenn and Gray 1999[91]
Results presented as cost savings per patient based on delay in disease progression although no detailed results summarising delay in disease progression are provided. Reports that the analysis indicates healthcare cost savings may be produced that to some extent would offset therapy costs.

Mild AD:
The greatest cost savings per patient are made over a two-year time horizon when £1227 is saved. Over a one-year period cost savings are only £85 and for 26 weeks (length of the trial of efficacy) only £10 is saved per patient.

Moderate AD:
Over the longest time horizon of two years cost savings are £777 per patient, less than when treating the mild AD group. However, greater savings are achieved by treating patients with moderate AD, over the shorter time horizons when £356 is saved over one year and £48 is saved over 26 weeks.

Sensitivity analysis: none

Hauber *et al.*, 2000[92]
Results presented as estimates of the delay in cognitive decline and the cost savings attributable to this delay for time horizons of six months, one year and two years. Reports that although rivastigmine may result in cost savings for moderately demented patients, the most cost-effective indication may be for patients in the early stages of AD. This conclusion is dependent on the life expectancy of the patients.

Mild AD:
The model predicts treatment will delay progression to the moderate stages of AD by 4 days, 25 days or 56 days for the six-month, one-year and two-year time horizons respectively. Once these patients reach a stage of moderate AD, progression to severe AD is predicted to be delayed by an additional one day over one year or 69 days over two years. There was no additional delay over a six-month period. The total delay in progression to severe AD over a two-year period is therefore predicted to be 125 days. Cost savings (independent of drug treatment cost) were estimated to be US$132, US$836, US$4839 over the six-month, one-year and two-year time horizons (two-year cost savings discounted).

Moderate AD:
The model predicts treatment will delay progression to severe AD by three days, 22 days or 51 days over six months, one year and two years respectively. Cost savings (independent of drug treatment cost) were estimated to be US$137, US$980, US$2290 over six months, one year and two years (two-year cost savings discounted).
The cost savings for the average patient in the trial population were also presented and estimated to be US$134, US$907, US$3578 over the six-month, one-year and two-year time horizons respectively.

Sensitivity analysis: none

continued

Hauber *et al.* 2000[93]

Results presented as estimates of the delay in cognitive decline and the cost savings attributable to this delay in cognitive decline, over time horizons of six months, one year and two years. A threshold analysis to estimate the QALY gains required for treatment to be cost-effective is also presented using two assumptions for QALY value. Reports that treatment with rivastigmine clearly yields savings in the cost of caring for AD patients that, on average, exceed the costs of the drug after 2 years of treatment. When taking each disease stage separately however, direct cost savings only exceed drug costs for patients with mild AD after two years of treatment.

Mild AD:

The model predicts treatment will delay progression to give an additional four days, 33 days or 188 days in the mild stage for the six-month, one-year and two-year time horizons respectively. Average daily cost savings (independent of drug treatment cost) were estimated to be Can $0.45, Can $1.85, Can $6.44 over the six-month, one-year and two-year time horizons (two-year cost savings discounted). Average net cost of treatment was only cost saving over the two-year time horizon (–1538 Can$). For the shorter time horizons treatment was not cost saving.

The QALY gains required for treatment to be cost effective were at the most 0.0466 (with a QALY valued at $20,000). No QALY gains were required at two years for either of the QALY values tested.

Mild to moderate AD:

Treatment was estimated to give an additional six days, 43 days or 106 days in the mild-to-moderate stage for the six-month, one-year and two-year time horizons respectively. Average daily cost savings (independent of drug treatment cost) were estimated to be Can $0.83, Can $3.34, Can $4.13 over the six-month, one-year and two-year time horizons (two-year cost savings discounted). Treatment was not cost saving over any of the time horizons; estimates of the average net cost of treatment ranged from Can $652 to 147 for six months and two years respectively. The QALY gains required for treatment to be cost effective ranged from 0.0015 to 0.0326.

Moderate AD:

Treatment was estimated to give an additional eight days, 33 days or 44 days in the moderate stage for the six-month, one-year and two-year time horizons respectively. Average daily cost savings (independent of drug treatment cost) were estimated to be Can$ 1.29, Can$ 2.74, Can$ 1.82 over the six-month, one-year and two-year time horizons (two-year cost savings discounted). Treatment was not cost saving with drug costs included over any of the time horizons; cost estimates ranged from Can $568 to 1,834 for six months and two years respectively. The QALY gains required for treatment to be cost effective ranged from 0.0060 to 0.0917.

All stages:

Days saved ranged from five to 137 accompanied by average daily cost savings of Can $0.71 to 4.93 from the six-month to two-year time horizons. Average net cost of treatment was cost saving for the two-year time horizon (Can $432). No QALY gains were required for treatment to be cost effective at two years; for the shorter time periods the maximum QALY gain required was 0.0346 (with a QALY valued at $20,000).

Sensitivity analysis: none

Summary results: economic evaluations for galantamine

Getsios et *al*.[95]
Mild to moderate AD:
Patients in the galantamine group predicted to remain in pre-FTC for 5.3% longer, spending 9.9% less time requiring FTC.
The model assumes no survival advantage.
A QALY gain of 0.05 years (0.04 discounted) is predicted for all patients.
Delaying time to FTC is expected to save $4910 (CDN) in health care costs, where the cost of galantamine is excluded.
When galantamine costs are included in the analysis a cost saving of $788 (CDN) is expected.

Moderate AD:
Patients in the galantamine group predicted to remain in pre-FTC for 10.1% longer, spending 11.2% less time requiring FTC.
The model assumes no survival advantage.
A QALY gain of 0.07 years (0.06 discounted) is predicted for patients with moderate AD.
For the subgroup of patients with moderate AD the cost savings are expected to be $7097 (CDN) excluding galantamine costs, and $3718 (CDN) where galantamine costs are included.

Sensitivity analysis (mild to moderate AD):
Authors report that the most important cost inputs are the cost of galantamine and cost of FTC.
Cost of galantamine + 50%: cost saving alters to additional cost of $1273 (CDN), with cost per discounted QALY at $30,042 (CDN).
Predicted cost savings are eliminated where the cost of galantamine increased by 19%.
Cost of FTC + 50%: cost saving alters to additional cost of $1845 (CDN)
Predicted cost savings are eliminated where the cost of FTC is at 15% below the base-case estimate.
The proportion of FTC in nursing homes is also an important input.

Garfield et *al*.[96]
Mild to moderate AD:
Patients in the galantamine group predicted to see a reduction of 10% in time requiring FTC at home (a reduction of 9.9% in FTC in an institution).
Expected net cost saving of SEK27,436 (EUR 3131) over period of analysis.

Moderate AD:
Patients in the galantamine group predicted to see a reduction of 11.4% in time requiring FTC at home (a reduction of 11.2% in FTC in an institution).
Expected net cost saving of SEK49,019 (EUR 5594) over period of analysis.

Sensitivity analysis (mild to moderate AD):
Authors report that univariate analysis found net cost savings predicted with galantamine persisted in most scenarios.
In sensitivity analysis on price of galantamine the predicted cost savings were eliminated when costs are 91% higher than base estimated price.
In Sweden the proportion of patients requiring FTC who are in nursing home or psychogeriatric ward is estimated to be 85%. Sensitivity analysis reports that for every 20% change in this proportion net costs change by SEK 4724 (EUR 539). Where the proportion is assumed to be zero a net saving of SEK 7346 (EUR 838) is predicted.
[we are unsure if the sensitivity analysis here includes cost for galantamine]
Nursing home care was an important cost driver. Predicted cost savings were eliminated where cost is 39% lower than base estimate.
Where survival effects were included patients treated with galantamine were predicted to live an average of 1.16 months longer than untreated patients. In this instance costs savings are predicted to disappear as significant additional costs are incurred for patients surviving this extended period of time.

Caro et *al*.[97]
Mild to moderate AD:
11% (discounted) reduction in time spent in FTC for patients treated with galantamine.
The mean amount of FTC avoided per patient treated is 0.144 years (discounted).
The mean QALY gained per patient treated reported at 0.039 (discounted)
Mean cost per patient over time on no drug treatment is reported at NLG 130,317, compared to NLG 127,267 in patients treated with galantamine (the majority of cost for both groups is institutional care). [cost saving of NLG 3,050]
The model assumes no survival advantage.

Sensitivity analysis (mild to moderate AD):
One-way sensitivity analysis reported. Authors report that sensitivity analysis indicates that it requires substantial departures from the main estimates for estimated cost savings to be eliminated. If no FTC patients are institutionalised, then net (additional) cost of NLG 5438 is predicted.

continued

Migliaccio-Walle *et al*.[98]
Mild to moderate AD:
In the base-case analysis, patients survived a mean of 59.9 months.
Estimated that patients are treated on galantamine for a mean of 32.8 months.
Estimated a reduction in the mean time in FTC of 2.6 months (16-mg dose) and 3.1 months (24-mg dose). Patients on 16-mg dose remained in pre-FTC for 7% longer and spent an estimated 12% less time in FTC. On 24-mg dose patients remained in pre-FTC for 8% longer and spent an estimated 14% less time in FTC.
Cost savings of between $2408 and $3601 per patient were predicted.
The cost of FTC in a nursing home was the largest cost component, contributing approx. 66% of the cost.
Galantamine treatment predicted to be economically dominant (greater effect and less cost), therefore cost-effectiveness ratios not reported.

Sensitivity analysis (mild to moderate AD):
Analyses undertaken using different time horizons to the 10-year base-case time horizon. Over 1 year the cost to avoid 1 discounted month of FTC with galantamine 16-mg was $13,384 ($9567 with 24 mg daily); beyond 3 years cost savings were predicted for both dose regimens.
Sensitivity analysis examined the effect of increasing the proportion of patients assumed to discontinue annually – with 16-mg daily treatment the predicted cost savings were eliminated when 24.8% of patients discontinued, with 24-mg regimen 39.3% of discontinuations were required before predicted cost savings were lost (baseline input parameter for discontinuation not clearly stated in the paper).
In sensitivity analysis a survival effect was introduced which resulted in increased survival of 1.2 months for 16-mg regimen (1.4 months on 24 mg), with an additional cost of $669 and an estimated $7674 per discounted life-year gained ($646 per life-year gained on 24 mg).
Sensitivity analysis considered treatment effect only in responders, using improvements in ADAS-cog, with much larger cost savings seen in these patients groups compared to the base-case analysis.
Authors report that other key inputs did not alter the direction of results.

Ward *et al*.[99]
Mild to moderate AD:
Galantamine 16-mg estimated to require FTC for 12% less time, and to delay the need for FTC by 2.5 months. Expected 10-year cumulative cost per patient at £28,615. A net cost increase of £481.
Galantamine 24-mg estimated to require FTC for 15% less time, and to delay the need for FTC by 3.02 months. Expected 10-year cumulative cost per patient at £28,806. A net cost increase of £672.
Treatment with galantamine predicted to increase the annual costs for the first 3 years, with costs in subsequent years predicted to be partially offset by delaying the need for FTC.
A mean gain of 0.06 QALYs is predicted (over 10 years) [QALY gains not presented by dose].
The model assumes no survival advantage.
Galantamine 16-mg results presented as equivalent to £8,693 per QALY [this does not calculate from the data presented above].

Moderate AD:
Small cost savings predicted at £228 [other results for these patients not explicitly stated].

Sensitivity analysis (mild to moderate AD):
Reducing the proportion of patients needing FTC who were admitted to an institution from 48% to 40% resulted in a net cost per patient (16 mg) of £731 a reduction to 35% resulted in a net cost per patient (16 mg) of £886.
Varying the utility estimate ± 50% resulted in the estimated cost per QALY for galantamine 16-mg ranging from £5,810 to £17,431.

Results reported to be non-sensitive to changes in discount rate.

Summary results: economic evaluations for memantine

Jones et al.[104]
Base-case analysis: Treatment with memantine (compared to no pharmacological treatment), over 2 years, was associated with an improvement of 0.10 years (SD: 0.04) in time spent in an independent state, a delay of 0.06 years (3 weeks) (SD: 0.04) before institutionalisation, an increase of 0.04 QALYs (SD: 0.04), and a cost reduction of £1963 over 2 years (SD: £4,504).

Sub-group analysis:
Was undertaken for four groups of patients, (a) initially moderately severe and independent (b) initially moderately severe and dependent (c) initially severe and independent, and (d) initially severe and dependent. In the first three of these subgroups improvements in clinical effect and cost savings were greater than the base-case analysis. In the latter group (d), memantine was associated with an improvement of 0.02 years (7 days) in time spent in an independent state, a delay of 0.04 years (14 days) before institutionalisation, an increase of 0.01 QALYs, and an incremental cost of £42 over 2 years.

Sensitivity analysis:
Sensitivity analysis is reported in relation to duration of treatment and efficacy.
The worst case scenario is presented as an improvement of 0.07 years in time spent in an independent state, 0.04 years in the community, an additional 0.03 QALYs and cost savings of £529.

François et *al.*[103]
Base case analysis:
Treatment with memantine (compared to no pharmacological treatment), over 5 years, was associated with an improvement of 0.34 years in time spent in an independent state, a delay of 0.08 years (approx. 1 month) before institutionalisation, and a cost reduction of €1,687.

Sensitivity analysis:
Probabilistic sensitivity analysis showed that in 93.7% of cases (base-case assumptions) treatment with memantine was cost saving. The authors report that where decision makers are willing to pay €30,000 per year of independency gain, there is a 99.7% probability of the intervention being cost-effective. Where treatment was assumed to be effective for only 6 months (compared to a base case of 1 year) authors report memantine treatment showed 2.5 additional months of independence, and delayed institutionalisation by 20 days, but with a cost reduction of only €61 over 5 years. Authors report that results were generally robust to other one-way sensitivity analysis undertaken. Where analysis was undertaken with no weightings for dependency level, authors report memantine as cost saving at €300.

Note:
Studies published only in abstract form provide limited information on methods and results, but Antonanzas *et al.*[102] and Guilhaume *et al.*[100] report findings from modelling studies (similar in design to those described above) that predict cost savings over time and benefits in terms of increasing time spent in an independent health state.

Appendix 15

Critical appraisal of cost-effectiveness models submitted to NICE by manufacturers/sponsors

The methodology used to critically appraise economic models submitted to NICE by industry is from Philips *et al*. (2004) (York University),[109] who have reviewed available evidence on economic modelling and its critical appraisal, and synthesised this in the context of HTA providing guidance on good practice and a framework for the critical appraisal of cost-effectiveness models.

Below we provide detail on the components of the critical appraisal format, i.e. dimensions of quality and key critical appraisal questions; thereafter we provide an outline appraisal of the cost-effectiveness model used in the industry submission to the NICE appraisal process.

Dimension of quality		Questions for critical appraisal
Structure		
S1	Statement of decision problem/objective	Is there a clear statement of the decision problem? Is the objective of the evaluation and model specified and consistent with the stated decision problem? Is the primary decision maker specified?
S2	Statement of scope/perspective	Is the perspective of the model stated clearly? Are the model inputs consistent with the stated perspective? Has the scope of the model been stated and justified? Are the outcomes of the model consistent with the perspective, scope and overall objective of the model?
S3	Rationale for structure	Has the evidence regarding the model structure been described? Is the structure of the model consistent with a coherent theory of the health condition under evaluation? Have any competing theories regarding model structure been considered? Are the sources of data used to develop the structure of the model specified? Are the causal relationships described by the model structure justified appropriately?
S4	Structural assumptions	Are the structural assumptions transparent and justified? Are the structural assumptions reasonable given the overall objective, perspective and scope of the model?
S5	Strategies/comparators	Is there a clear definition of the options under evaluation? Have all feasible and practical options been evaluated? Is there justification for the exclusion of feasible options?
S6	Model type	Is the chosen model type appropriate given the decision problem and specified causal relationships within the model?
S7	Time horizon	Is the time horizon of the model sufficient to reflect all important differences between options? Is the time horizon of the model, the duration of treatment and the duration of treatment effect described and justified? Has a lifetime horizon been used; if not has a shorter time horizon been justified?
S8	Disease states/pathways	Do the disease states (state transition model) or the pathways (decision tree model) reflect the underlying biological process of the disease in question and the impact of interventions?
S9	Cycle length	Is the cycle length defined and justified in terms of the natural history of disease?

continued

Dimension of quality	Questions for critical appraisal
Data	
D1 Data identification	Are the data identification methods transparent and appropriate given the objectives of the model? Where choices have been made between data sources, are these justified appropriately? Has particular attention been paid to identifying data for the important parameters in the model? Has the process of selecting key parameters been justified and systematic methods used to identify the most appropriate data? Have the quality of the data been assessed appropriately? Where expert opinion has been used, are the methods described and justified?
D2 Pre-model data analysis	Are the data analysis (pre-model) methodology based on justifiable statistical and epidemiological techniques?
D2a Baseline data	Is the choice of baseline data described and justified? Are transition probabilities calculated appropriately? Has a half-cycle correction been applied to both cost and outcome? If not, has this omission been justified?
D2b Treatment effects	If relative treatment effects have been derived from trial data, have they been synthesised using appropriate techniques? Have the methods and assumptions used to extrapolate short terms results to final outcomes been document and justified? Have alternative assumptions been explored through sensitivity analysis? Have assumptions regarding the continuing effect of treatment once treatment is complete been documented and justified? Have alternative assumptions been explored through sensitivity analysis?
D2c Costs	Are the costs incorporated into the model justified? Has the source for all costs been described? Have discount rates been described and justified given the target decision maker?
D2d Quality of life weights (utilities)	Are the utilities incorporated into the model appropriate? Is the source for the utility weights referenced? Are the methods of derivation for the utility weights justified?
D3 Data incorporation	Have all data incorporated into the model been described and referenced in sufficient detail? Has the use of mutually inconsistent data been justified (i.e. are assumptions and choices appropriate)? Is the process of data incorporation transparent? If data have been incorporated as distributions, has the choice of distribution for each parameter been described and justified? If data have been incorporated as distributions, is it clear that second order uncertainty is reflected?
D4 Assessment of uncertainty	Have the four principal types of uncertainty been addressed? If not, has the omission of particular forms of uncertainty been justified?
D4a Methodological	Have methodological uncertainties been addressed by running alternative versions of the model with different methodological assumptions?
D4b Structural	Is there evidence that structural uncertainties have been addressed via sensitivity analysis?
D4c Heterogeneity	Has heterogeneity been dealt with by running the model separately for different sub-groups?
D4d Parameter	Are the methods of assessment of parameter uncertainty appropriate? Has probabilistic sensitivity analysis been done; if not has this been justified? If data are incorporated as point estimates, are the ranges used for sensitivity analysis stated clear and justified?
Consistency	
C1 Internal consistency	Is there evidence that the mathematical logic of the model has been tested thoroughly before use?
C2 External consistency	Are the conclusions valid given the data presented? Are any counterintuitive results from the model explained and justified? If the model has been calibrated against independent data, have any differences been explained and justified? Have the results of the model been compared with those of previous models and any differences in results explained?

Donepezil: industry submission (cost-effectiveness) from Eisai/Pfizer

Dimension of quality		Appraisal
Structure		
S1	Statement of decision problem/objective	✓ Decision problem stated: economic impact of donepezil treatment for AD; the objective to compare expected costs and benefits of treatment of patients in mild to moderate AD with usual care or usual care plus donepezil using an economic model is consistent with this. NICE is the primary decision maker.
S2	Statement of scope/perspective	✓ Perspective of model clearly stated as that of the NHS, PSS and patient; patient costs are not included in the base-case analysis. Model inputs are generally consistent with the perspective (patient costs are included in sensitivity analysis). However, cost structure used makes no allowance for the fact that some institutionalised AD patients are self funding (not all costs fall on NHS and PSS). Scope of the model stated and justification provided. Outcomes consistent with perspective, scope and objective of the model.
S3	Rationale for structure	? Evidence for model structure is not described. No competing theories for model structure are discussed. Some data sources are specified but no evidence or justification is provided to indicate that changes in cognitive outcome measures (as reported in trials) correlate with disease progression in the model and the disease states defined by MMSE score. A growing evidence base suggests that cognition alone is not the best indication of disease progression (see report).
S4	Structural assumptions	? Structural assumptions are not always transparent and when stated, justification is not always provided, e.g. mortality. The suitability of the MMSE cognitive outcome measure (which was a secondary outcome measure in the clinical trial) as a marker of disease progression is not discussed/justified, and the use of MMSE as some marker of disease progression is a weakness in the model structure.
S5	Strategies/comparators	× Only drug treatment with donepezil is considered; other available drugs are not discussed – although this is consistent with the stated objective.
S6	Model type	? This type of state transition (Markov type) model is frequently used to assess this type of decision problem; however, in common with other published research, justification for the causal relationships is not provided. Transition probability data is derived from clinical trial outcome measures and is used to plot disease progression characterised by MMSE categories which describe AD severity groups.
S7	Time horizon	?/✓ A time horizon of 5 years is stated together with treatment duration and duration of treatment effect. A full justification for these choices is not provided. A lifetime time horizon is not used, the chosen time horizon is in line with that of a previous modelling approach. It is likely that the time horizon is long enough to reflect the differences between options. Treatment duration is based on previous NICE guidance; however, duration of effect is driven by relatively short-term clinical trial findings. Longer term effectiveness data would be preferable.
S8	Disease states/pathways	? The AD severity groups do reflect accepted categories of disease severity. However, the sole use of MMSE is a weakness in the modelling of AD progression. It is unclear whether transition probabilities obtained from one clinical trial (n = 286) accurately describe the impact of the intervention (generalisability of trial data is not discussed/justified).
S9	Cycle length	× A cycle length of 12 months is defined but this may be too long and no justification for this length of cycle is given. Results reported with a half-cycle correction differ markedly from those with no half-cycle correction.
Data		
D1	Data identification	?/× Data identification methods are not described and where alternative data sources are available, e.g. other clinical trials, the choices made have not been thoroughly justified. Cost data used were based on a UK sample/survey; alternative sources of cost data were tested in sensitivity analyses. The cost structure for AD, by severity, is not justified and costs used may overestimate treatment costs. No allowance has been made in cost data for the fact that not all institutional costs fall on the NHS and PSS budget. If quality assessment of the data took place this is not reported.
D2	Pre-model data analysis	? Pre-model data analysis included the calculation of transition probabilities which is incompletely described. Indirect data analysis/modelling has been undertaken to determine the cost estimates, and no discussion on this is provided.

continued

Dimension of quality	Appraisal
D2a Baseline data	? One clinical trial used as the source for the data to generate transition probabilities but no justification/support is given. The static transition probabilities employed in the base case appear to have been calculated appropriately, although the differences in the treatment and placebo probabilities appear large in the context of the effects reported in the clinical trial. Transit probabilities are distributed using the Dirichlet method when probabilistic sensitivity analysis is undertaken; it was not possible to check the accuracy of these methods. Half-cycle corrections are applied to costs and outcomes in the base case but it is not clear if a correction is applied to the results of the sensitivity analyses.
D2b Treatment effects	? No effectiveness parameters from the clinical trial are given; only the transition probabilities (calculated from changes in MMSE scores) are reported. The model extrapolates short-term results to final outcomes assuming donepezil benefits begin to decline after the first year of treatment such that after 5 years donepezil-treated patients are at the same average MMSE level as untreated patients. Two alternative scenarios regarding duration of treatment effect are tested in sensitivity analyses. Justification for continuing clinical benefits of donepezil beyond the 1-year trial period comes from observational studies (no discussion/support provided). The generalisability of trial data is not discussed.
D2c Costs	See discussion in S2, D1 & D2 above for costs. Discounting described and justified.
D2d Quality of life weights (utilities)	The model does not incorporate quality of life data; the authors state that the difficulties associated with estimating quality of life have necessitated the use of pragmatic endpoints in their AD model.
D3 Data incorporation	?/✓ Data incorporated into the model are referenced and in general are described well. The process of data incorporation is transparent and it is clear when point estimates or distributions have been used, although in some cases where data have been incorporated as distributions; e.g. transition probabilities for non-severe states, the choice of distribution and justification of this choice is not well described.
D4 Assessment of uncertainty	? Parameter uncertainty was the only type of uncertainty addressed. The other three areas of uncertainty were not addressed as indicated below.
D4a Methodological	?/✓ The model was run with different methodological assumptions – half-cycle correction (limited reporting), effectiveness scenarios, input values. We would have liked to have seen some attempt to consider other endpoints (e.g. QALYs)
D4b Structural	?/✓ (see above/below)
D4c Heterogeneity	× The authors chose not to run the model for different sub-groups of AD patients, but did state that this was because they preferred instead to estimate an average cost-effectiveness.
D4d Parameter	? An attempt has been made to investigate parameter uncertainty using probabilistic sensitivity analysis (PSA) for three variables (NHS and social costs, drug costs, and transition probabilities) and one-way sensitivity analysis for another four variables (discount rate, disease progression, usual care costs, and mortality rate). It is not clear why only some variables were assessed by PSA and the methods by which PSA of transition probabilities was carried out are unclear.
Consistency	
C1 Internal consistency	× No evidence is presented or referenced to indicate that the mathematical logic of the model had been tested.
C2 External consistency	The conclusion is drawn that donepezil costs can be fully offset (i.e. cost saving), generally after 2 years of treatment but the results presented are for a 5-year period with cost savings only being indicated in some of the sensitivity analyses. Where cost savings are reported it is unclear whether a half-cycle correction has been employed, and if not whether the results hold when this is introduced. The model was not calibrated against independent data. The results of the model were compared with those of previous models, although not in detail, and found to be more favourable. Possible reasons for the differences, e.g. use of different cost sources and 12-month transition probabilities, were commented on with the authors asserting their belief that the PSA they conducted for costs and transition probabilities gave a more robust assessment than previous studies which used point estimates of these parameters. The authors also point out that the probabilistic base-case results are similar to the deterministic results.

Rivastigmine: industry submission (cost-effectiveness) from Novartis

Dimension of quality		Appraisal
Structure		
S1	Statement of decision problem/objective	✓ Decision problem stated: the cost effectiveness of rivastigmine; the objective to compare expected costs and benefits of treatment of patients in AD with rivastigmine versus no treatment, using an economic model is consistent with this. The primary decision maker is not explicitly stated but appears to be NICE.
S2	Statement of scope/perspective	✓ Perspective of model stated as that of the NHS and social services, so patient and carer costs are omitted. Model inputs are generally consistent with the perspective, although cost inputs take no account of the fact that a proportion of patients are self funding in institutional care. Scope of the model stated and justification provided. Outcomes consistent with perspective, scope and objective of the model.
S3	Rationale for structure	? Evidence for model structure is not described. No competing theories for model structure are discussed. The paper describing the natural history model used to model baseline change in MMSE over time is cited[135] but not summarised, whilst the data source describing the open-label rivastigmine trial outcomes is unpublished and not summarised. It is therefore not possible to determine whether the changes in cognitive outcome measures found in the trial correlate with disease progression in the model and the disease states defined by MMSE score. A growing evidence base suggests that cognition alone is not the best indication of disease progression (see report).
S4	Structural assumptions	? A number of structural assumptions are stated with limited justification provided. × Structural assumption over relationship between utility scores and the MMSE score is specified, but given the methods employed, and the basis for bias, we highlight a concern with this assumption/relationship. × Baseline AD progression used a method published by Mendiondo and colleagues,[128] using MMSE scores. This methodology is based on a US study, but no discussion is provided on the generalisability of findings to the UK. The suitability of the MMSE cognitive outcome measure as a marker of disease progression is not discussed, and current literature indicates that cognitive function alone is not appropriate to model AD progression (see report).
S5	Strategies/comparators	✓ Drug treatment with rivastigmine is considered; this is consistent with the stated objective.
S6	Model type	? A Markov-type model is used based on MMSE and includes the definition of stages of disease severity (for cost inputs) defined using MMSE. The Markov-type model is appropriate, but the use of MMSE to drive the model is not appropriate; it is used owing to the nature of the clinical trial data available. The model uses a natural history-type model developed by Mendiondo[128] in combination with data from an open-label, non-randomised and non-controlled study. The appropriateness of these methods is uncertain; we note that we have not been able to check the equations used to predict disease progression as they are not presented in the reporting of the model. Justification for the causal relationship between MMSE scores and disease progression is not provided, and current literature indicates it may be inappropriate to use cognitive function alone to model AD progression (see report).
S7	Time horizon	?/✓ A time horizon of 5 years is used, and this seems appropriate for AD progression in an elderly cohort of patients. Duration of treatment effect is based on trial data, but an open label study is used that may be subject to selection bias (not ITT analysis, responders only).
S8	Disease states/pathways	× AD progression is based solely on MMSE score, and as above we believe this to be inappropriate.
S9	Cycle length	?/✓ A cycle length of 6 months is used; no justification for this length of cycle is given, but we feel it is appropriate in the context of AD.
Data		
D1	Data identification	? Data identification methods are not described in any detail, but where alternative data sources are available some justification is provided for the choices made with alternatives tested in sensitivity analyses. The cost data chosen were felt by the authors to be the most robust resources and were based on the UK population (the submission does

continued

Dimension of quality	Appraisal
	provide a narrative review on the cost literature, but not a systematic review). The utility values associated with MMSE scores were estimated by a mapping process that involved a number of assumptions and interpretations, and as such the relationship between MMSE and utility score remains uncertain, but despite the opportunities for measurement error alternatives for this parameter do not seem to have been widely explored. If quality assessment of the data took place this is not reported.
D2 Pre-model data analysis	? Pre-model data analysis of baseline data used to obtain the expected disease progression of the patients that would have occurred had they not been receiving rivastigmine (specific equations not provided). This data analysis is based on a published methodology, but validation of the method and justification are not presented by the model authors. ? Pre-model analysis was also used to obtain a regression function for the utility–MMSE relationship. This is described in detail, but validity and appropriateness are uncertain, with subjective judgements and opportunity for bias (as below)
D2a Baseline data	? The methods used to model AD progression, baseline and treatment cohorts are inappropriate due to the reliance on MMSE alone. The progression of AD over time in the treatment group is inappropriate. The method used to model baseline AD progression is not presented in full, and there is no discussion supporting the generalisability of the methodology to the UK patient group. A half cycle correction does not appear to have been used, except in the case of dropouts, which are assumed to drop out midway through each six-month cycle of the model.
D2b Treatment effects	× No effectiveness parameters from the unpublished open-label trial are reported. Equations relating change in MMSE score to disease progression are used to determine what the course of AD would have been if patients had not been receiving rivastigmine treatment. The effectiveness is based on an open label study, analysis is not intent-to-treat, and the effect is based on a comparison with treatment responders (those continuing treatment). Treatment effects are more pronounced in later years when the majority of patients had stopped treatment.
D2c Costs	See discussion on costs in S2, D1 & D2. Discounting described and justified.
D2d Quality of life weights (utilities)	?/× Quality of life weights incorporated into the model were derived by mapping health state data from clinical trials onto the Health Utilities Index, version III (HUI III). The mapping exercise and establishment of a relationship between utility and MMSE score involved a number of assumptions, interpretation between different measurement scales, and data from various sources. Therefore, given the opportunities for measurement error and the range of uncertainties within the quality of life estimates, SHTAC feel the findings should be regarded as illustrative/experimental.
D3 Data incorporation	?/✓ Data incorporated into the model are referenced and in general are described well with the exception **[commercial/academic confidential information removed]**. The process of data incorporation is transparent and it is clear when point estimates or distributions have been used and in general justification for the choices made is provided.
D4 Assessment of uncertainty	? Parameter uncertainty was the only type of uncertainty addressed. The other three areas of uncertainty were not addressed as indicated below.
D4a Methodological	× The model was not run with different methodological assumptions (e.g. effectiveness measures, AD progression methods, QALY methods).
D4b Structural	X Alternative model structures have not been tested, although different model inputs are used in sensitivity analysis.
D4c Heterogeneity	X The authors did not run the model for different sub-groups of AD patients.
D4d Parameter	?/✓ Parameter uncertainty was investigated using probabilistic sensitivity analysis (PSA) for eight variables (death rate, clinical pathways, mean MMSE, probability of institutionalisation, cost in institutions, costs at home, monitoring costs and utility MMSE regression coefficients) and one-way sensitivity analysis of included costs, death rate, probability of institutionalisation, change in institutional and home costs, and change in monitoring costs. × Methods used to undertake probabilistic analysis appear incorrect, with a random selection between various mean values used for a number of parameters, rather than a distribution around a selected mean parameter value.

continued

Dimension of quality		Appraisal
Consistency		
C1	Internal consistency	× No evidence is presented or referenced to indicate that the mathematical logic of the model had been tested.
C2	External consistency	The conclusion is drawn that rivastigmine offsets close to half of its drug costs through saved expenditure after the first six months of treatment but this result is not presented; only the results for a 5-year period are reported. It is unclear whether a half-cycle correction has been employed, and if not whether the results hold when this is introduced. The model was not calibrated against independent data. The results of the model are not compared with those of previous models.

Galantamine: industry submission (cost-effectiveness) from Shire Pharmaceuticals (based predominantly on published study by Ward *et al.*[99])

Dimension of quality		Appraisal
Structure		
S1	Statement of decision problem/objective	✓ Decision problem stated clearly – to assess the long-term health and economic impact of treating mild to moderate Alzheimer's disease with galantamine (16 mg or 24 mg per day) compared to no cholinesterase therapy. Using an economic model is consistent with this. The primary decision maker is not explicitly stated but appears to be NICE.
S2	Statement of scope/perspective	✓ Perspective of model clearly stated as that of the NHS and PSS. Time spent caring by informal caregivers and caregiver distress was analysed and reported separately. Model inputs are consistent with the perspective (although cost data may not be true NHS and PSS cost). Scope of the model stated and justification provided. Outcomes consistent with perspective, scope and objective of the model.
S3	Rationale for structure	? Evidence for model structure is described by Caro *et al*. (2001).[131] The structure is based on the probability of needing FTC, which is predicted by a range of patient characteristics. Unlike some other models, disease progression in this model is not described solely by changes in cognitive outcome measures, although differences between treatment and placebo cohort are due to difference in cognitive function and presence of psychotic symptoms. Caro *et al*. explain the benefits of their modelling approach in comparison with other existing theories for model structure. Sources of evidence used to develop and inform the model are specified.
S4	Structural assumptions	? Structural assumptions remain uncertain, and there is a heavy reliance on the use of predictive risk equations to model 'need for FTC' and 'death' in the patient cohort. Mortality in the model appears an underestimate of that expected in a UK patient cohort. In general, accepting the prediction of disease progression is adequate, the approach seems reasonable and in line with the stated objective, perspective and scope of the model.
S5	Strategies/comparators	✓ Drug treatment with galantamine is considered; this is consistent with the stated objective.
S6	Model type	?/✓ This model has been developed specifically to address the health economics of Alzheimer's disease based on the need for full time care, an outcome that can be applied across a range of populations and settings.
S7	Time horizon	X/? A time horizon of 10 years is used. We believe that 10 years may be too long for the model time horizon, as many patients will die within 5–6 years, and many patients will be in the FTC state for the majority of the time period. A 5-year time horizon may be more appropriate.

continued

Dimension of quality		Appraisal
S8	Disease states/pathways	? This model does not involve states of differing AD severity; instead the model predicts when patients will reach a state of FTC. The characteristics predicting disease progression came from a study of 236 US patients; the applicability of these factors to other populations is not discussed. The model does reflect the progressive nature of AD and the increased need for care in the more severe stages of the disease. However, it may be that the two AD states are too crude to reflect the underlying biological process of AD, although they do focus on important health policy endpoints. The impact of the intervention is seen as a delay in the need for FTC.
S9	Cycle length	? The model considers/presents disease progression over 1-year periods, with predictive risk equations for shorter time periods used in pre-model analysis.
Data		
D1	Data identification	?/× Data identification methods are not described and there is very little discussion of alternative data sources. Data sources for resource use and costs are from UK sources. If quality assessment of the data took place this is not reported.
D2	Pre-model data analysis	? Pre-model analysis was used by Caro *et al.*[131] to obtain the hazard functions for this model. Limited information is provided with the model on the data analysis, but citations are provided to appropriate methodological references.
D2a	Baseline data	? The cohort data used in the cited Caro study[131] to identify the predictors of disease progression for which hazard functions were calculated came from an American study of 236 patients followed for 7 years. The hazard functions reported by Caro were then used in the industry submission model to simulate the course of disease progression in a hypothetical cohort of patients under assumptions of treatment and no treatment. The authors give no indication of whether the natural history of AD in the US cohort is likely to be the same in another population and therefore whether this model is applicable to the UK. ? Data on resource use from two UK national surveys (from 1980s); sources cited but no justification provided ? Data on utilities from a published source (USA sample); no discussion of justification provided.
D2b	Treatment effects	? Effectiveness data were not summarised in the description of the model although it was stated that the efficacy measure was the change in ADAS-cog and three RCTs were cited (Raskind *et al.*, Wilcock *et al.*, Tariot *et al.*). However, ADAS-cog values had to be converted to mMMS scores for entry into the model and this involved several steps which may have introduced uncertainty and measurement error. The methods used to extrapolate the short term trial results are described in the cited publications by Ward *et al.*[99] and Caro *et al.*[131] Alternative assumptions have not been explored in sensitivity analyses.
D2c	Costs	See discussion on costs in S2, D1 & D2a. Discounting described and justified.
D2d	Quality of life weights (utilities)	?/× Quality of life weights are incorporated into the model which were derived from published data[ref Neumann]. The methods used to derive the utility weights are not reported in detail. It appears that assumptions have been made about where pre-FTC and FTC states lie in relation to severities of mild, moderate, severe and profound AD but a full justification for these choices is not provided.
D3	Data incorporation	?/× Data incorporated into the model are referenced but some, particularly the effectiveness data, are not described in sufficient detail. It is not clear when point estimates or distributions have been used.
D4	Assessment of uncertainty	× Parameter uncertainty and heterogeneity were addressed to a limited extent by Ward *et al.*[99] but this was not reported in the industry submission to NICE. Other areas of uncertainty were not addressed as indicated below.
D4a	Methodological	× The model was not run with different methodological assumptions.
D4b	Structural	× Alternative model structures have not been tested
D4c	Heterogeneity	?/✓ Ward *et al.*[99] used the model for a moderate subgroup of patients and for a sub-group defined as those responding to galantamine (with improved or maintained cognition) after 6 months and who continue treatment.

continued

Dimension of quality		Appraisal
D4d	Parameter	×/? One-way sensitivity analyses only (for key input parameters including the proportion of patients in institutional care, the cost per month and utilities and one-way sensitivity analysis for the discount rates for costs and benefits.
Consistency		
C1	Internal consistency	✓ Evidence is referenced by Caro *et al.*[131] to indicate that the mathematical logic and validity of the model has been tested.
C2	External consistency	? The conclusion is drawn that galantamine therapy is predicted to increase the annual costs of care for the first 3 years of treatment, but costs in subsequent years will be offset by delaying the need for FTC; however only the results for the full 10-year period are presented. As would be expected, more substantial savings are predicted for the subset of patients who respond to galantamine in the first 6 months and continue treatment. The model was not calibrated against independent data. There have not been any previous models of galantamine cost-effectiveness with which these results could have been compared.

Memantine: industry submission from Lundbeck

Dimension of quality		Appraisal
Structure		
S1	Statement of decision problem/objective	✓ Decision problems stated: 1) to estimate the incremental cost-effectiveness of memantine vs standard care over 2 years in patients with moderately severe to severe AD; 2) to assess the incremental cost-effectiveness of memantine in the specific population of patients with moderately severe to severe AD receiving a stable dose of donepezil. NICE is the primary decision maker.
S2	Statement of scope/perspective	✓/? 1) Perspective of model stated as NHS and PSS, but there is some uncertainty over the consistency of cost inputs with this perspective. Cost estimates are high, and do not make allowance for the fact that not all institutional costs for AD are met by the NHS and PSS. ✓/? 2) **[Commercial/academic confidential information removed]**
S3	Rationale for structure	?/✓ 1) Some evidence regarding the model structure has been described. The assumptions regarding the duration of efficacy are based on results of a 6-month, double-blind, clinical trial, and an open label study that began at the end of this trial and continued for 6 months. We have some concerns over the data used to model disease progression over time, and the data used to adjust transit probabilities by treatment strategy. ?/✓ 2) **[Commercial/academic confidential information removed]**
S4	Structural assumptions	×/?/✓ 1) Structural assumptions related to definition of health states by severity, location and dependency are reasonable. However, MMSE score (cognitive function alone) is used as a measure of severity, and this is a weakness with the model. The use of calculated odds ratios to adjust disease progression over time is questionable, based on the data used and the explanation provided.
S5	Strategies/comparators	Drug treatment with memantine is considered, as is drug treatment in patients already receiving stabilised doses of donepezil. No other available drugs are discussed.
S6	Model type	✓ A Markov model was used to estimate cost-effectiveness of memantine compared to standard care. The stages of the model were defined according to severity, dependency (dependent or independent) and setting (community or institution). × 2) **[Commercial/academic confidential information removed]**
S7	Time horizon	✓ 1) The time horizon is stated at 2 years. Justification is provided, claiming to take into account that outcomes were dependent on a chronic progressive disease state and the low life expectancy among moderately severe to severe AD patients. × 2) **[Commercial/academic confidential information removed]**

continued

Dimension of quality		Appraisal
S8	Disease states/pathways	?/✓ 1) The disease states appear to reflect the underlying biological process of AD. 13 states are included 3 (moderate, moderately severe, severe) x 2 (institution, community) x 2 (independent, dependent) + 1 (death). As above, we have concerns over the use of MMSE alone to define disease severity. ?/✓ 2) **[Commercial/academic confidential information removed]**
S9	Cycle length	✓ 1) The cycle length (6 months) is defined. Justification is provided – the length of the cycle was based on clinical practice, consistent with the clinical trials and followed NICE guidance of assessing AD patients every 6 months.
Data		
D1	Data identification	× 1+2) Data identification methods have not been described. There does not appear to have been a systematic method of identification of the most appropriate data. It is not clear whether the parameters for the model were set and then an appropriate data source found or whether the parameters were set to match the available data source. There is no evidence of quality assessment. Data used in the model (costs, utilities) do not appear to reflect the perspective of the study, and may not be from sources that are generalisable to the UK patient group.
D2	Pre-model data analysis	×/? Transit probabilities are from various sources, with rates used remaining unpublished. Data on transit probabilities may not be valid and generalisable to the UK patient population. Data and methods are not transparent. ×/? Odds ratios are calculated to adjust model transit probability rates for dependency and location, with methodology not fully reported and justified.
D2a	Baseline data	? 1) Baseline data have been based on a large-scale observational/longitudinal study, where epidemiological, clinical and costs data were collected (LASER-AD Study). This study, non-comparative and non-controlled, is open to bias (e.g. selection bias) and the sample may not be representative and generalisable to the UK treatment population. The data on transition probabilities are presented, but the validity (power of the trial) to determine transit probabilities is questionable. The process for handling uncertainty with respect to transit probabilities – probabilistic analysis using a Dirichlet distribution – is described. Half-cycle corrections do not appear to have been applied to costs and outcomes. No half-cycle correction has been applied to cost or outcome (this would appear acceptable with a 6-month cycle in a chronic slowly progressing disease like AD). 2) **[Commercial/academic confidential information removed]**
D2c	Costs	See discussion on costs in S2, D1 & D2a. Discounting at 3.5% for future costs & benefits. This reflects future NICE references case methods.
D2b	Treatment effects	× 1) No effectiveness parameters are reported from cited clinical trials. One RCT is cited, but the effectiveness measures do not relate directly to the transit probabilities used. Based on guidelines, the treatment duration was considered to be 18 months, as efficacy was considered to last 12 months. The total 12 months of drug efficacy was based upon the results of the 6-month, double-blind clinical trial and continued for 6 months. After 1 year, transition probabilities in the memantine group were those from the standard care strategy or no pharmacological treatment.
D2d	Quality of life weights (utilities)	X/? 1) Utility weights are reported and referenced (although data remain unpublished). We have serious concerns over the methods and the generalisability of the utility estimates used. ×/? 2) **[Commercial/academic confidential information removed]**
D3	Data incorporation	As previous comments, concerns over incorporation of data inputs.
D4	Assessment of uncertainty	The assessment of uncertainty has been addressed in some cases, but not in others. See below for further details.
D4a	Methodological	× Alternative versions of the models with different methodological assumptions do not appear to have been carried out.
D4b	Structural	✓ 1) Structural uncertainties appear to have been addressed with the use of probabilistic sensitivity analyses and one-way sensitivity analyses.
D4c	Heterogeneity	✓ 1) Heterogeneity has been considered with four different subgroup analyses. × 2) **[Commercial/academic confidential information removed]**

continued

Dimension of quality	Appraisal
D4d Parameter	✓ 1) Parameter uncertainty has been considered. A probabilistic sensitivity analysis was constructed to control for most uncertainties in the model. The model used a stochastic approach (Monte Carlo simulation) to take account of variability in each parameter through the use of *a priori* distributions, and to increase the confidence applied to the results. ✓ 2) **[Commercial/academic confidential information removed]**
Consistency	
C1 Internal consistency	×/? Some evidence to indicate that the mathematical logic of the models has been tested.
C2 External consistency	External validity is uncertain. ×/? 1) The conclusions made are valid given the data presented; however, we have raised serious concerns above on data inputs to the model. ✓37/? 2) **[Commercial/academic confidential information removed]**

Appendix 16

Figures showing the progression of AD over time using the AHEAD predictive risk equation for FTC, where no mortality effect is present in the SHTAC model

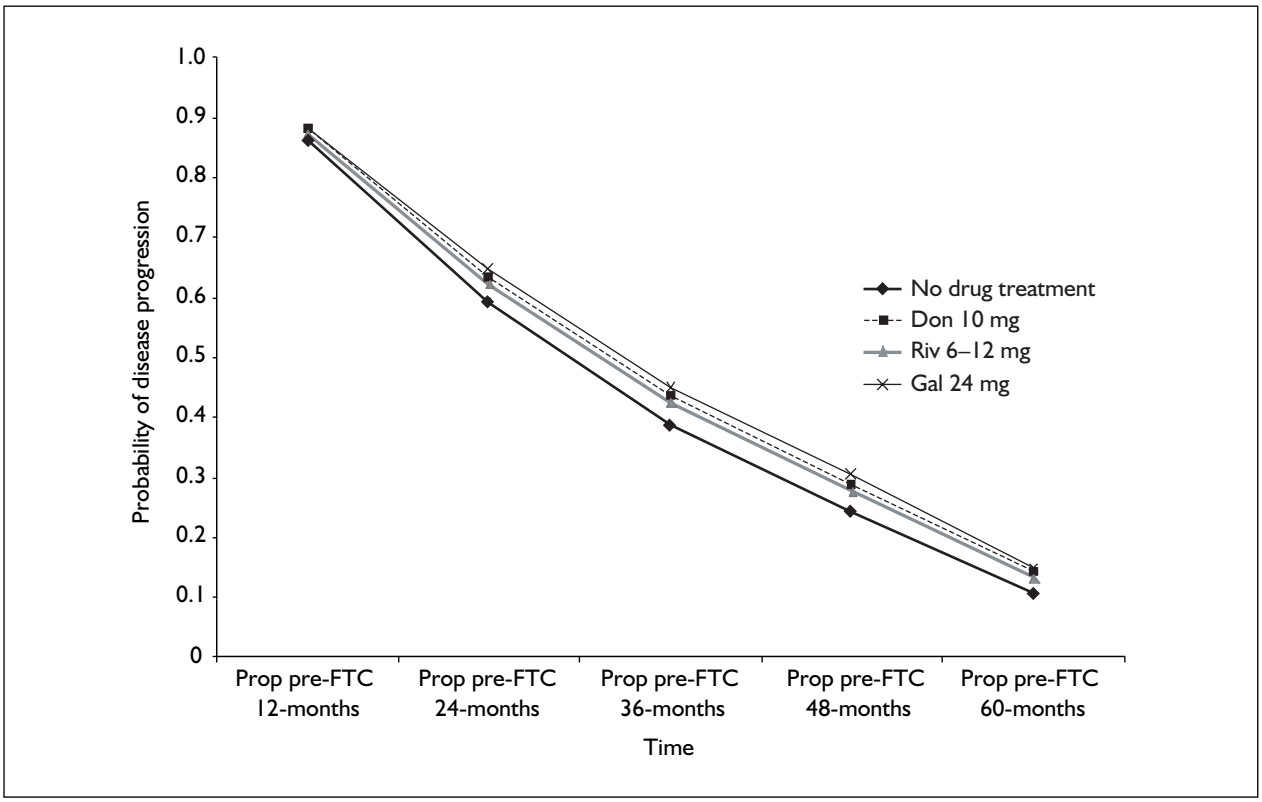

FIGURE 24 *Projected probabilities for patients remaining in pre-FTC over years 1–5, SHTAC model with no mortality impact in the model*

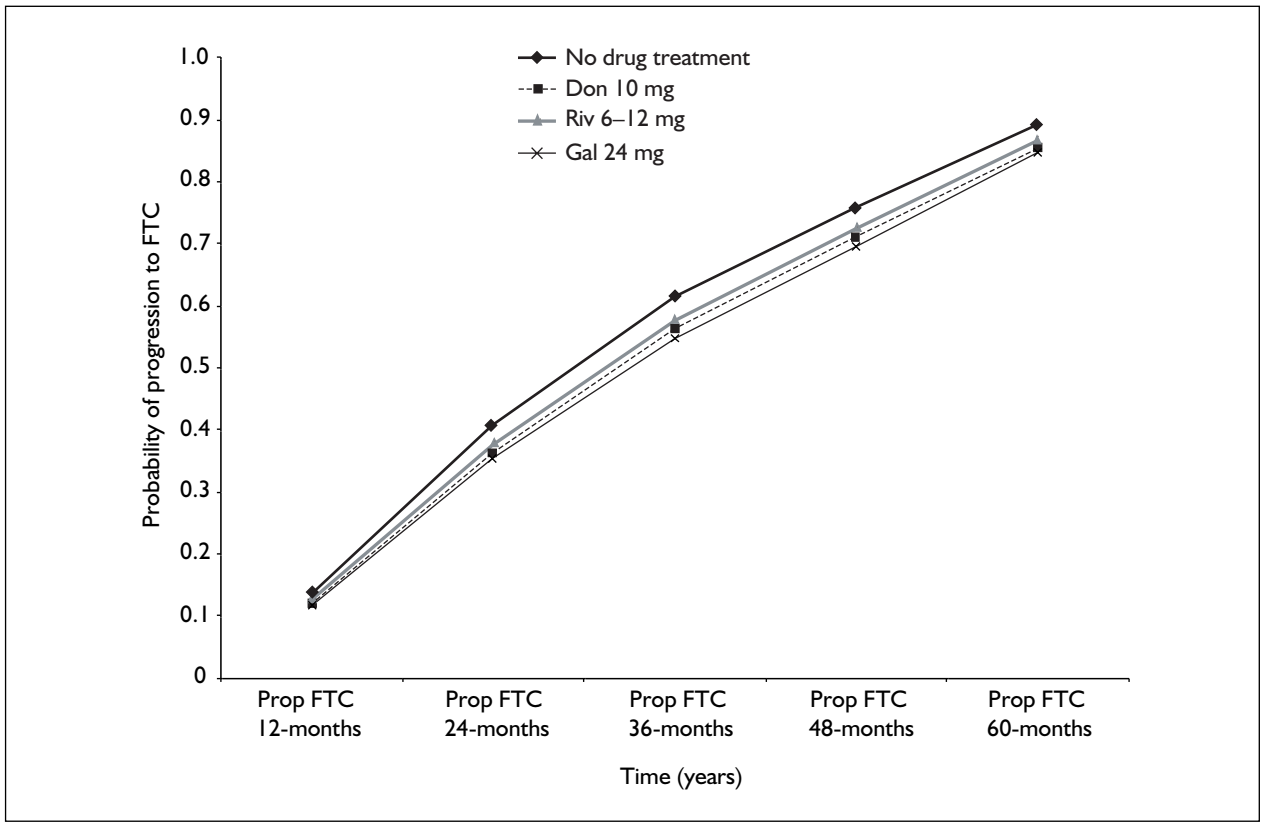

FIGURE 25 *Projected probabilities for patients progressing to FTC over years 1–5, SHTAC model with no mortality impact in the model*

Appendix 17

Resource use data from OPCS survey of elderly with cognitive disability

[Commercial/academic confidential information removed]

Appendix 18

Sensitivity analysis, against deterministic results, selected one-way and multiple sensitivity analysis using the model with parameter/input variations

Results of sensitivity/inputs	Cost per QALY		
	Donepezil 10 mg	**Rivastigmine 6–12 mg**	**Galantamine 24 mg**
Base case	£80,941	£57,985	£68,042
No discounting	£80,211	£56,841	£67,084
Discount rates at 3.5% for future costs and outcomes	£85,858	£61,234	£72,036
Assume all institutional costs met by NHS and PSS	£71,736	£48,781	£58,838
Assume 85% of institutional costs met by NHS and PSS	£76,339	£53,383	£63,440
Assume presence of EPS and psychotic symptoms in 20% of cohort	£73,391	£52,037	£61,359
Assume a benefit of treatment on psychotic symptoms: with 20% reduction in the proportion of patients showing presence of symptoms at:			
Base case (10% with symptoms)	£76,215	£54,367	£64,240
20% with psychotic symptoms	£70,102	£49,637	£59,139
75% with psychotic symptoms	£45,593	£30,583	£38,359
AD Duration (mean):			
2.66 years (SD 1 year)	£86,512	£62,369	£72,964
Assumptions on location of FTC:			
20% community 80% inst.	£66,979	£44,024	£54,082
60% community 40% inst.	£71,532	£61,476	£71,532
Assume no entry cost (drug/monitoring) for treatment cohort	£61,691	£42,910	£51,196
Assume all costs met by NHS and 80% of FTC patients in institution	£51,638	£28,683	£38,743
Assume additional health care costs for institutional patients at (per year):			
£1000	£87,376	£64,420	£74,476
£10,000	£72,426	£49,471	£59,528
Assume costs for pre-FTC are (per month/cycle):			
£150 (SD £75)	£73,546	£50,590	£60,647
£600 (SD £300)	£92,233	£69,277	£79,332
Assumptions on health state utilities:			
Pre-FTC 0.75 FTC 0.20 (diff = 0.55)	£38,263	£27,411	£32,165
Pre-FTC 0.50 FTC 0.20 (diff = 0.30)	£70,149	£50,254	£58,970
Pre-FTC 0.50 FTC 0.34 (diff = 0.16)	£131,529	£94,226	£110,568
Effectiveness estimate:			
+1 point on base-case ADAS-cog	£56,108	£38,885	£47,728
– 1 point on base-case ADAS-cog	£130,587	£95,533	£106,262

continued

Results of sensitivity/inputs	Cost per QALY		
	Donepezil 10 mg	**Rivastigmine 6–12 mg**	**Galantamine 24 mg**
Mortality at (per year):			
0%	£62,175	£43,317	£51,608
8% – all patients	£74,887	£53,252	£62,739
15.4% – all patients	£89,916	£65,003	£75,904
Assume all costs met by NHS, 80% of FTC patients in institution, and a mortality annual rate of 8% (all patients)	£45,640	£24,006	£33,496
Assume additional (to base case) monitoring cost:			
+ 1 additional outpatient visit per year	£89,274	£66,140	£75,731
+ 1 outpatient appointment, and an additional 2 GP visits per year	£93,286	£70,067	£79,433
Note: GP visit assumed to cost £26 (assume SD of £13), PSSRU.			

Feedback

The HTA Programme and the authors would like to know your views about this report.

The Correspondence Page on the HTA website (http://www.ncchta.org) is a convenient way to publish your comments. If you prefer, you can send your comments to the address below, telling us whether you would like us to transfer them to the website.

We look forward to hearing from you.

The National Coordinating Centre for Health Technology Assessment,
Mailpoint 728, Boldrewood,
University of Southampton,
Southampton, SO16 7PX, UK.
Fax: +44 (0) 23 8059 5639 Email: hta@soton.ac.uk
http://www.ncchta.org

ISSN 1366-5278