The effectiveness and cost-effectiveness of donepezil, galantamine, rivastigmine and memantine for the treatment of Alzheimer's disease (review of TA111): a systematic review and economic model

Produced by: Peninsula Technology Assessment Group (PenTAG), University of Exeter

Authors:

Mary Bond	Research Fellow in HTA	PenTAG, University of Exeter
Gabriel Rogers	Research Fellow in HTA	PenTAG, University of Exeter
Jaime Peters	Research Fellow in Modelling	PenTAG, University of Exeter
Rob Anderson	Senior Lecturer in Health Economics	PenTAG, University of Exeter
Martin Hoyle	Research Fellow in Modelling	PenTAG, University of Exeter
Alec Miners	Lecturer in Health Economics	London School of Hygiene and Tropical Medicine
Tiffany Moxham	Information Specialist	PenTAG, University of Exeter
Sarah Davis	Senior Lecturer in Health Economics	University of Sheffield
Praveen Thokala	Research Associate	University of Sheffield
Allan Wailoo	Director of the Decision Support Unit	University of Sheffield
Mike Jeffreys	Consultant Physician	Royal Devon and Exeter Foundation Trust Hospital
Chris Hyde	Professor of Public Health	PenTAG, University of Exeter

Correspondence to:
Mary Bond
PenTAG, Peninsula Medical School
Veysey Building
Salmon Pool Lane
Exeter EX2 4SG

Competing interests: None

Date completed: June 18th 2010

About the Peninsula Technology Assessment Group (PenTAG)

The Peninsula Technology Assessment Group is part of the Institute of Health Service Research at the Peninsula College of Medicine & Dentistry. PenTAG was established in 2000 and currently has four major work streams; independent Health Technology Assessments for NICE and the NIHR HTA Programme, systematic reviews and economic analyses for the NICE Centre for Public Health Excellence, systematic reviews as part of the Cochrane Collaboration Heart Group and evidence synthesis work in relation to the needs of the SW Peninsula Collaboration for Applied Health Research and Care (PenCLAHRC), as well as for other local and national decision-makers.

The group is multi-disciplinary and draws on individuals' backgrounds in public health, health services research, computing and decision analysis, systematic reviewing, statistics and health economics. The Peninsula College of Medicine & Dentistry is a school within the Universities of Plymouth and Exeter. The Institute of Health Research is made up of discrete but methodologically related research groups, among which Health Technology Assessment is a strong and recurring theme. Recent projects include:

Health Technology Assessment

Dasatinib and nilotinib for imatinib resistant or intolerant chronic myeloid leukaemia.

Bevacizumab, sorafenib tosylate, sunitinib and temsirolimus for renal cell carcinoma: A systematic review and economic evaluation.

Systematic review of the effectiveness and cost-effectiveness of weight management schemes for the under fives.

The effectiveness and cost-effectiveness of cochlear implants for severe to profound deafness in children and adults: a systematic review and economic model.

The effectiveness and cost-effectiveness of methods of storing donated kidneys from deceased donors: a systematic review and economic model.

The harmful health effects of recreational ecstasy: a systematic review of observational evidence.

The use of surrogate outcomes in model-based cost-effectiveness analyses: a survey of UK health technology assessment reports.

Systematic review and economic analysis of the comparative effectiveness of different inhaled corticosteroids and their usage with long-acting beta2 agonists for the treatment of chronic asthma in children under the age of 12 years.

The Clinical Effectiveness and Cost-effectiveness of Cardiac Resynchronisation (Biventricular Pacing) for Heart Failure: a systematic review and economic model.

Synthesising Public Health Evidence

Prevention of unintentional injuries to children in outdoor play and leisure environments

Legislation, regulations, standards, enforcement and strategic policies for the prevention of unintentional injuries to children

Front matter

Source of funding: This report was commissioned by the NIHR HTA Programme as project number 09/87/01.

Declared competing interests of the authors: None

Acknowledgements: We would like to thank: Dr Jane Wolstenholme (Health Economics Research Centre, Oxford), the principal investigators of the Oxfordshire Alzheimer's disease cohort study for sharing their study's dataset; Prof. Gill Livingston (UCL) for sharing the LASER-AD study dataset with us, Prof. Douglas Galasko (University of California San Diego; developer of the ADCS-ADL and the ADCS-ADL-sev functional status instruments) Researchers at the Personal Social Services Research Unit, University of Kent (in relation to clarifying aspects of the costs in the Unit Costs of Health and Social Care 2009 report) and Prof. Douglas Galasko and Dr Steven Stokes from the University of California, San Diego for help with the use of the ADCS-ADL questionnaires.

We would also like to acknowledge the help of, Martin Pitt for model checking, and Mark Pearson and Harriet Hunt for supporting the systematic reviews, Louise Crathorne for proof reading, Will Stahl-Timmins for summary graphics, Colin Green for general support in helping us to understand this complex disease and modelling area and Sue Whiffin and Jenny Lowe for their administrative support all from the Peninsula Medical School, University of Exeter.

Expert Advisory Group

We would particularly like to thank Prof. Peter Passmore who acted as our Expert Advisor for his help throughout the project.

Competing interests Prof. Peter Passmore has received fees for consultancy, honoraria and assistance with study leave from Pfizer, Eisai, Shire, Jansen and Jansen, Novartis and Lundbeck and has given expert testimony in relation to the last Judicial Review as an expert for Eisai.

Rider on responsibility of the report: The views expressed in this report are those of the authors and not necessarily those of the NIHR HTA Programme. Any errors are the responsibility of the authors.

Contributions of authors

Mary Bond	Provided overall project management, wrote the protocol, assessed abstracts and titles for inclusion, contributed to the clinical effectiveness systematic review, contributed to the design of the model and the writing and editing of the report.
Gabriel Rogers	Assessed abstracts and titles for inclusion, led the systematic review of clinical effectiveness, contributed to the design and execution of the economic model and contributed to the writing and editing of the report.
Jaime Peters	Led the design, development and execution of the economic model and contributed to the writing and editing of the report.
Rob Anderson	Oversaw the cost-effectiveness aspects of the analysis and report, advised on obtaining costs and utilities for the model, contributed to the design of the model and contributed to the editing of the report.
Martin Hoyle	Contributed to the design, parameterisation of the model and model checking and contributed to the editing of the report.
Alec Miners	Critically appraised industry submissions and contributed to the editing of the report.
Tiffany Moxham	Wrote and ran all the search strategies.
Sarah Davis	Provided critical appraisal of the cost-effectiveness model submitted by Eisai and contributed to the writing of the final report.
Praveen Thokala	Provided critical appraisal of the cost-effectiveness model submitted by Eisai and contributed to the writing of the final report.
Allan Wailoo	Provided critical appraisal of the cost-effectiveness model submitted by Eisai and contributed to the writing of the final report.
Mike Jeffreys	Provided clinical input into the design of the model, advised on clinical matters and contributed to the editing of the report.
Chris Hyde	Led the systematic review of economic evaluations, contributed to the design of the model, contributed to the writing and editing of the report and was overall Director of the project and Guarantor of the report.

Table of contents

List of tables

List of figures

List of abbreviations

AChEI	Acetylcholinesterase inhibitors
AD	Alzheimer's Disease
ADAS-cog	Alzheimer's Disease Assessment Scale cognitive subscale.
ADCS-ADL	Alzheimer's Disease Cooperative Study Activities of Daily Living inventory
ADFACS	Alzheimer's Disease Functional Assessment and Change Scale
ADL	Uniform Activities of Daily Living
AEs	Adverse Events
AHEAD	Assessment of Health Economics in Alzheimer's disease
BADLS	Bristol Activities of Daily Living Scale
BDS	Blessed Dementia Scale
BEHAVE-AD	Behaviour Pathology in Alzheimer's Disease
BGP	Behavioural Rating Scale for Geriatric Patients
BOI	Burden of Illness
BPSD	Behavioural and psychological symptoms of dementia
BSC	Best Supportive Care
BVRT	Benton Visual Retention Test
CCOHTA	Canadian Coordinating Office for Health Technology Assessment
CDN	Canadian
CDR	Clinical Dementia Rating Scale
CDR-SB	Clinical Dementia Rating Scale – Sum of the Boxes
CE	Cost-effectiveness
CEA	Cost-effectiveness analysis
CGIC	Clinical Global Impression of Change
CGRS	Crichton Geriatric Rating Scale
CHEC	City Health Economic Centre
CI	Confidence Interval
CiC	Commerical in Confidence
CIBIC	Clinician's Interview-based Impression of Change
CMCS	Caregiver-rated Modified Crichton Scale
CRD	Centre for Reviews and Dissemination
CSRI	Client Service Receipt Inventory
CUA	Cost–utility analysis

DAD	Disability Assessment for Dementia scale
DES	Discrete Event Simulation
DH	Department of Health
DLB	Dementia Lewy Body
DOMINO-AD	Donepezil and Memantine in Moderate-to-Severe Alzheimer's Disease
DSM-4	Diagnostic and Statistical Manual of mental disorders
DSS	Department of Social Services
DSST	Digit Symbol Substitution Test
DSU	Decision Support Unit
EOAD	Early Onset Alzheimer's Disease
EPS	Extra Pyramidal Symptoms
EQ-5D	EuroQol-5 Dimensions
FAST	Functioning Assessment Staging Scale
FDA	Food and Drug Administration
FHSA	Family Health Service Authority
FLD	Frontal Lobe Dementia
FOME	Fuld Object Memory Evaluation
FTC	Full Time Care
GAS	Global Attainment Scale
GBS	Gottries-Brane-Steen scale
GDS	Global Deterioration Scale
GHQ	General Health Questionnaire
HRQL	Health-Related Quality of Life
HTA	Health Technology Assessment
HUI	Health Utilities Index (versions II or III)
IADL	Instrumental Activities of Daily Living
ICD-10	International Classification of Diseases-10
ICER	Incremental cost-effectiveness ratio
IDDD	Interview for Deterioration in Daily living in Dementia
IHQL	Index of Health Related Quality of Life
IQWiG	Institute for Quality and Efficiency in Healthcare
ISPOR	International Society for Pharmacoeconomics and Outcomes Research
ITT	Intention to treat analysis
LOCF	Last observation carried forward
MCMC	Markov Chain Monte-Carlo
MENFIS	Mental Function Impairment Scale

MMSE	Mini-mental state examination
MDS	Minimum Data Set
MRI	Magnetic resonance imaging
MTA	Multiple Technology Appraisal
MTC	Mixed Treatment Comparison
n	number
NADES	National Dementia Economic Study
NHS	National Health Service
NHS CRD	NHS Centre for Reviews and Dissemination
NICE	National Institute for Clinical Excellence
NICE-SCIE	National Institute for Clinical Excellence Social Care Institute for Excellence
NINCDS-ADRDA	National Institute of Neurology and Communicative Disorders and Stroke – Alzheimer's Disease and Related Disorders Association
NMDA	n-methyl-D-aspartate
NNT	Number Needed to Treat
NOSGP	Nurse Observation Scale for Geriatric Patients
NPI	Neuropsychiatric Inventory
NPI-D	Neuropsychiatric Inventory - Distress Subscale
NS	Not statistically significant
NYU	New York University
N/A	Not assessed
OC	Observed Case(s)
ONS	Office for National Statistics
OPCS	Office of Population Census and Surveys
p	Probability
PCMD	Peninsula College of Medicine and Dentistry
PCT	Primary Care Trust
PD	Progressive Disease
PDS	Progressive Deterioration Scale
PGA	Patient Global Assessment Scale
PRC	Prolonged Release Capsule
PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-analyses
PSMS	Physical Self-Maintenance Scale
PSS	Personal Social Services
PSSRU	Personal Social Services Research Unit

QALY	Quality-adjusted life year
QoL	Quality of Life
RCT	Randomised controlled trial
SEK	Swedish Krona
SEM	Standard Error of Mean
SEVINT	OPCS measure of intellectual functioning
SHTAC	Southampton Health Technology Assessment Centre
SIB	Severe Impairment Battery
SMD	Standardised Mean Difference
SSD	Social Services Department
TAR	Technology Assessment Report
TESS	Treatment Emergent Signs and Symptoms
TMT	Trail making test
UPDRS	Unified Parkinson's Disease Rating Scale
USD	US Dollars
VAD	Vascular Dementia
WIHRD	Wessex Institute for Health Research and Development
WMD	Weighted mean difference

1. Summary

1.1. Background

Alzheimer's disease is the most commonly occurring form of dementia, accounting for approximately 62% of instances of this condition. Alzheimer's disease is predominantly a disease of later life with 5% of the UK population over 65 years affected. Early onset Alzheimer's disease can be found in younger people, this is a rare condition, only accounting for an estimated 2.2% of those with dementia. Alzheimer's disease is more commonly found in women than men in the UK, with 67% of women with dementia having Alzheimer's disease but only 55% of men. However, the association with gender is completely explained by the shorter life-expectancy of men.

The incidence of dementia therefore increases with age. In England and Wales, for people aged 65 to 69 years, the incidence is estimated to be 7.4 (95% CI 3.6-16.1) per 1,000 person years, this rises to 84.9 (95% CI 63.0-107.8) per 1,000 person years at 85 years old and above. These rates predict 180,000 new cases of dementia per year and, if 62% of these have Alzheimer's disease (see above) then there are approximately 111,600 new cases in England and Wales per year. The Medical Research Council's Cognitive Function and Aging Study (2006) found, that in England and Wales, increasing age was the greatest risk factor for dementia, with gender weakly associated. Having Parkinson's disease increased the risk of dementia by three times, odds ratio 3.5 (95% CI 1.3-9.3), but rating your own health as poor was a greater risk factor, odds ratio 3.9 (95% CI 2.2-6.9). Better education was a marginally protective factor, 0.7 (95% CI 0.5-1.0).

It is generally believed that the causes of Alzheimer's disease are multi-factoral, with increasing age bringing the greatest risk. Up to 5% of cases are linked to genetic causes; medical history and lifestyle are also contributing factors.

There is currently no cure for Alzheimer's disease. The time taken from diagnosis to death varies; an estimated median survival for Alzheimer's disease from onset has been calculated as 7.1 years (95% CI 6.7-7.5 years) in the USA by Fitzpatrick and colleagues and is reported in Warrell and colleagues as about 10 years in the UK. The difference may depend on

whether survival figures are calculated from time of reported onset or time of actual diagnosis; in general a diagnosis of Alzheimer's disease halves life-expectancy.

The contribution of Alzheimer's disease to these survival figures is difficult to know, as people with Alzheimer's disease frequently have co-morbidities which will influence their longevity. The proportion of deaths estimated to be due to Alzheimer's disease increases with age and varies with gender.

Interventions

This technology assessment report considered four pharmaceutical interventions. Three have marketing authorisations in the UK for the treatment of adults with mild to moderately severe Alzheimer's disease (measured by the MMSE 26-10). These are donepezil (Aricept®, manufactured by Eisai), rivastigmine (Exelon®, manufactured by Novartis), and galantamine (Reminyl®, manufactured by Shire Pharma). They are acetylcholinesterase inhibitors (AChEI), which work by restricting the cholinesterase enzyme from breaking down acetylcholine, thus increasing the concentration and duration of acetylcholine at sites of neurotransmission. The fourth drug, memantine hydrochloride (Ebixa®) manufactured by Lundbeck, has a UK marketing authorisation for the treatment of people with moderate to severe Alzheimer's disease (measured by the MMSE, score of 20 or less). It is a voltage-dependent, moderate-affinity, uncompetitive N-methyl-D-aspartate (NMDA) receptor antagonist that blocks the effects of pathologically elevated toxic levels of glutamate that may lead to neuronal dysfunction.

Comparators

The comparators for these drugs are dependent on disease severity, classified by the Mini Mental State Examination (MMSE) criteria:

- mild AD (MMSE 21-26): donepezil; galantamine and rivastigmine
- moderate AD (MMSE 10-20: donepezil; galantamine, rivastigmine and memantine
- severe AD (MMSE <10): memantine

All of the above were also compared with best supportive care (i.e. without treatment with any AChEIs or memantine).

Population

The population is adults with Alzheimer's disease. However, as in the assessment which informed TA111, where trials have included participants with mixed dementias, these trials will be included where the dominant dementia is Alzheimer's disease.

Outcome measures

The outcomes of interest include measures of:

- Severity of disease and response to treatment
- Behavioural symptoms
- Mortality
- Ability to remain independent
- Likelihood of admission to residential/nursing care
- Health related quality of life of patients and carers (where data permit, analysis will be carried out separately for patients alone, and for patients and carers combined)
- Adverse effects of treatment
- Cost-effectiveness and costs (review of economic studies)

Study design

For the review of clinical effectiveness, only systematic reviews of randomised controlled trials (RCTs) and RCTs were considered. The review protocol made provision for broadening search criteria to include some observational evidence if insufficient systematic reviews or RCTs were identified; however, this proved unnecessary in view of the reasonable yield of evidence of a preferred design.

The inclusion and exclusion criteria for the systematic review of economic evaluations were identical to those for the systematic review of clinical effectiveness, except:

- Non-randomised studies were included (e.g. decision model based analyses, or analyses of patient-level cost and effectiveness data alongside observational studies.)
- Full cost-effectiveness analyses, cost–utility analyses, cost-benefit analyses and cost consequence analyses were included. (Economic evaluations which only report

average cost-effectiveness ratios will only be included if the incremental ratios can be easily calculated from the published data).

- Standalone cost analyses based in the UK NHS were also sought and appraised.

1.2. Methods

1.2.1. Clinical effectiveness systematic review

Data sources

Electronic databases were searched for systematic reviews and/or meta-analyses, randomised controlled trials (RCT) and ongoing research in November 2009 and updated in March 2010; this updated search revealed no new includable studies. The databases searched included: The Cochrane Library (2009 Issue 4, CDSR and CENTRAL), MEDLINE, MEDLINE In Process, Embase, PsycINFO, EconLIT, ISI Web of Science Databases: Science Citation Index, Conference Proceedings Citation Index- and Biosis, the CRD databases: NHSEED, HTA, and DARE databases. Where possible a controlled trials and human filter was added. As this is an update of a previous review, the searches were run in the timeframe 2004-current. The meta-register of controlled trials and clincaltrials.gov were searched for ongoing trials. Bibliographies of included studies were searched for further relevant studies. The reference lists of the industry submissions were also scrutinised for additional studies. Due to resource limitations the search was restricted to English language papers only. All references were managed using Reference Manager (Professional Edition Version 11; Thomson ISI ResearchSoft) and Microsoft Access 2003 software.

Study selection

Relevant studies were identified in two stages. Titles and abstracts returned by the search strategy were examined independently by two researchers (GR and MB) and screened for possible inclusion. Disagreements were resolved by discussion. Full texts of the identified studies were obtained. Two researchers (GR and MB) examined these independently for inclusion or exclusion, and disagreements were again resolved by discussion.

Data extraction

Data were extracted by GR and checked by MB. Disagreements were resolved by discussion.

Data synthesis

Where data permitted, the results of individual trials were pooled using the following methods:

- Pairwise meta-analysis: We used random-effects meta-analyses only, regardless of any statistical evidence of inter-study homogeneity. Heterogeneity was explored by visualisation of results and, in statistical terms, both Cochran's *Q* (compared to a χ^2 distribution) and the I^2 statistics. Small-study effects (including publication bias) were visualised using funnel plots and quantified using Egger's test.

- Pooling of multiple outcome measures: In addition to pairwise meta-analyses of treatment effect, pooled on each outcome's natural scale (weighted mean difference), we combined outcomes in a series of broad domains – cognitive, functional, behavioural, and global – to investigate the overall characteristics of reported effectiveness evidence in each area. In order to combine studies using different outcome measures within each domain, effect sizes were expressed as a standardised mean difference.

- Meta-regression: Where there was sufficient evidence (at least five individual datapoints in a meta-analysis), study level regression ("meta-regression") was used to explore the statistical heterogeneity across studies. Three prespecified covariates were explored: population age, population sex, and baseline disease severity (as measured by MMSE). Because of inconsistencies in the evidence-base, it was not possible to undertake multivariate analyses, so regressions were conducted solely on a univariate basis.

- Mixed treatment comparison – indirect comparison: In addition to pairwise meta-analyses, where sufficient data was available, we synthesised information on all technologies and their comparators simultaneously, in a mixed treatment comparison (MTC) using Bayesian Markov Chain Monte-Carlo (MCMC) sampling.

1.2.2. Review of past economic evaluations

A systematic review was conducted to update the systematic review of cost-effectiveness studies which was conducted in 2004 as part of the review of evidence to inform the NICE's earlier guidance on these drugs (TA111).

The review aimed to summarise the main results of the included studies, and identify any key economic costs and trade-offs relevant to the decision problem. It also indicated the strengths and weaknesses of different modelling approaches in this treatment area.

The search strategy included all those databases searched for clinical effectiveness and in addition NHSEED and Econlit. The inclusion and exclusion criteria for the systematic review of economic evaluations were identical to those for the systematic review of clinical effectiveness with the exception of the target study designs. The review targeted economic evaluations including decision model based analyses, analyses of patient-level cost and effectiveness data alongside RCTs and observational studies, full cost-effectiveness analyses, cost–utility analyses, cost-benefit analyses, cost consequence analyses and standalone cost analyses based in the UK NHS. Narrative synthesis, supported by the data extraction tables, was used to summarise the evidence base

1.2.3. Appraisal of industry submissions

Four manufacturer submissions were potentially available for this MTA. However, Novartis did not submit an economic evaluation and Shire only provided a critique of aspects of the previous SHTAC model which they felt remained unaddressed. The remaining two manufacturers both submitted economic evaluations based on decision models, which were both critiqued in depth. In addition, the decision support unit (DSU) was asked to examine the technical accuracy of the Eiasi / Pfizer economic evaluation for donepezil, as it was produced using software (ARENA) requiring specialist expertise.

1.2.4. PenTAG cost–utility model

An in-depth consideration and exploration of various modelling approaches and limited availability of data led to the development of a decision model based broadly on the structure of the three-state Markov model described in the previous technology assessment report; based upon time to institutionalization, parameterised with updated estimates of

effectiveness, costs and utilities. A review of all documentation (from manufacturers, interest groups, NICE and the published literature) relating to the decision model described in the previous TAR was undertaken to develop a list of key criticisms or perceived weaknesses of the previous model. Using this list, a number of changes to the model structure and the parameter values used in the three-state Markov model were implemented. The model was developed in MicroSoft Excel 2007 with additional analyses undertaken in the statistical software package R. A detailed description of the process undertaken to arrive at the PenTAG model can be found in Section 8.2.

For the three cholinesterase inhibitors (where rivastigmine capsules were considered separately to rivastigmine patches), the base-case analysis modelled a cohort of people with mild to moderate Alzheimer's disease. For memantine, the base-case analysis concerned people with moderate to severe Alzheimer's disease. In exploratory sensitivity analyses, the cost-effectiveness of treatment with donepezil, rivastigmine (capsules and patches) and galantamine was investigated for a cohort of people with mild Alzheimer's disease. Further exploratory sensitivity analyses investigated the cost–utility of donepezil, rivastigmine(capsules and patches), galantamine *and* memantine for people with moderate only Alzheimer's disease and the cost-effectiveness of memantine in the treatment of people with severe only Alzheimer's disease.

Disease progression based on age, MMSE and ADL was modelled using individual patient data from the UK-based study by Wolstenholme and colleagues (patient data from 1988 to 1999 in Oxfordshire). A prevalent cohort is assumed for this decision model since the data informing the parameter values are from a prevalent cohort. The study data supplied by Wolstenholme and colleagues also provided estimates of the NHS and PSS costs associated with Alzheimer's disease. Data from the LASER-AD longitudinal cohort study was also used to justify and/or corroborate a number of assumptions in the model.

The model starts when treatment begins for the treated cohorts. For the initial treatment period, mean time to institutionalization and mean time to death are predicted using mean baseline characteristics of the cohort. After the initial treatment period, any treatment effects are assumed to be observed, and so mean time to institutionalization is then predicted based on the mean baseline characteristics plus the mean absolute treatment effect for the treated cohorts. This leads to treated cohorts having a delay in institutionalization compared to best supportive care. The length of this treatment period was dependent upon the length of

follow-up reported in the source RCTs, and was defined to be six months. The PenTAG model allows for treatment discontinuations, and also assumes that for the three cholinesterase inhibitors, treatment stops once they enter institutionalization. Thus, the model is implicitly assuming that institutionalization is equivalent to severe Alzheimer's disease (MMSE < 10). Therefore, once in an institution, patients' quality of life and utility are assumed to be that of people with severe Alzheimer's disease (MMSE <10). For memantine, treatment is not assumed to stop once individuals are in institutional care, although there is also allowance for treatment discontinuations in the moderate to severe cohort. Both time to institutionalization and death are significantly dependent on age and therefore the cohort model allows three subgroups defined by age to be included, allowing us to model some degree of heterogeneity within the cohort.

A monthly time cycle was used in the model, and the time horizon was set at 20 years. By this time it was estimated that <5% of the cohort would be alive.

1.3. Clinical effectiveness results

1.3.1. Number and quality of effectiveness studies

From 1843 titles and abstracts screened four systematic reviews and 17 RCTs were found that matched our inclusion criteria that had been published since 2004. There were 12 pairwise comparisons with placebo (donepezil 5, n=234; galantamine 3, n=1386; rivastigmine 3, n=1995 and memantine 1, n=350); four head-to-head studies and one combination therapy study (memantine added to AChEIs); taken as a whole the quality of the trials was disappointing. A particular criticism is the use of LOCF and OC methods to account for missing data; these methods are inappropriate in a condition which naturally declines to death and may lead to an overestimation of the treatment effect. Methods of randomisation and allocation concealment were frequently not reported.

1.3.2. Summary of benefits and risks

In 2004 the assessment group found that donepezil improved cognitive and global outcomes, with increased benefit from higher doses, in some cases this benefit was maintained over a year. There was weaker evidence for a significant effect with functional and behavioural outcomes. The 2010 systematic review found five small poor quality studies which have

added to the evidence base. They had a maximum of six months follow-up. All studies measured cognitive outcomes. A dose related beneficial effect was found at 10 mg/day. One study measured functional and global outcomes but it was of such poor quality the positive findings lack credibility.

We found an additional three variable quality RCTs of galantamine v. placebo to add to the evidence base of six studies included in 2004. The previous review found a dose-response relationship for cognitive, functional and global outcomes. In the two trials reporting behavioural outcomes, one found a significant gain, the other did not. The studies included in our review all found significant benefit on cognitive outcomes; the results for functional and global outcomes were inconclusive, and no significantly positive gain was found for behavioural outcomes. However, when the results from these studies were pooled, significant gains for people taking galantamine were found for cognitive, functional and global outcomes.

The evidence for the effectiveness of rivastigmine in the previous review was varied; there was some evidence of benefit at 6-12mg/day with cognitive, functional and global outcomes, but no gain was reported on behavioural measures. Our update review found three more studies; one of these was of reasonable size and quality. Positive benefits from rivastigmine were found on cognitive, functional and global outcomes, but, as before, not on behavioural ones. The lower dose transdermal patch (9.5 mg/day) was shown to be as effective as the capsule (12 mg/day) but with fewer side effects.

There was some evidence, from a single study, in the previous review that memantine was more effective on cognitive and functional outcomes than placebo; although, as this study's results were not analysed by ITT, they may be unreliable. However, the new, poorer quality study, failed to show any benefit from memantine on any outcome measure. When the data were pooled, a significant benefit from memantine was found from global outcomes. It should be noted that these results are based on two moderate to poor quality trials and may be untrustworthy.

Three new head-to-head comparisons were found in addition to the three in the previous review. Only one of the new studies was large and of reasonable quality, this compared donepezil to rivastigmine. It measured cognitive, functional, behavioural and global outcomes, but only found statistically significant differences on functional and global outcomes, both favouring rivastigmine. This is in contrast to the much smaller and poorer

quality studies found in the previous review, which showed no significant differences between the treatments. One new study and one previous study compared donepezil with galantamine; neither were good quality. The trial from the previous review found that donepezil had greater effects on cognitive and functional outcomes. The new study only looked at global outcomes and found no difference between the treatments. One very poor quality study, looking at behavioural outcomes, compared all three AChE inhibitors; it found that rivastigmine was significantly better than donepezil or galantamine.

We also found one new, reasonably good, study comparing combined memantine with an AChE inhibitor against AChE inhibitor and placebo. This showed no significant advantage to combining these treatments. This contrasts with the results from the previous review which found significant benefits from combination therapy on cognitive, functional, behavioural and global outcomes. The reason for this difference in outcomes may be due to an underlying pharmacological interaction between galantamine and memantine - which neutralizes their respective effects - in the new trial, which used all three AChE inhibitors, whilst the existing trial only combined memantine with donepezil. The other difference between these studies is the lack of ITT analysis in the former one which may have led to more favourable results for combination therapy.

Overall we found that although more evidence has accumulated over the last six years, its impact on conclusions about effectiveness appears small. An enduring problem is that of trying to predict what will happen to people over the course of five years or more on the basis of six months or less information. It remains is impossible to say whether one AChE inhibitor is better than another at treating Alzheimer's disease. Important gaps in the evidence continue concerning long-term outcomes, impact on quality of life, carers and time to institutionalisation.

1.4. Cost-effectiveness results

1.4.1. Published economic evaluations

The systematic review of economic evaluations identified 23 included studies, over a third of which were only published as abstracts and could not be considered in depth. Of the remainder, most addressed the costs and cost-effectiveness of either donepezil or memantine. Of these, the majority reapplied modelling approaches considered as part of the

last guidance to the circumstances applying in other countries and were thus felt to add little to this update reconsidering cost-effectiveness in England and Wales. Enhanced modelling approaches were presented for both donepezil and memantine, but in both cases the publications closely mirrored the economic models submitted as part of the industry submissions.

1.4.2. Industry submissions

Two cost-effectiveness models were submitted and appraised: Eisai Ltd and Pfizer Ltd for donepezil and Lundbeck for memantine.

The model for donepezil has been described as a discrete event simulation model. This is a modelling approach which theoretically could overcome a number of challenges facing the assessment of the cost-effectiveness of drug treatments for AD, particularly dealing with multiple interdependent outcomes. However, the model does not employ a pure discrete event simulation approach and actually incorporates elements of individual sampling alongside some cohort modelling methods. The manufacturer's conclusion is that donepezil provides benefits at reduced costs relative to best supportive care, and is thus dominant, in both mild and moderately severe AD, a conclusion which is robust to the sensitivity analyses conducted by the manufacturer. However, the review of the submitted model identified several areas where there was concern with respects to the quality of the inputted data or the validity of the model assumptions. Exploratory sensitivity analyses examining plausible alternative assumptions suggest that the cost-effectiveness could be at the margins of what would normally be considered cost-effective by NICE.

The model for memantine used a more traditional Markov approach with three states, pre full time care, full time care and death. It concludes that memantine provides benefits at reduced costs relative to best supportive care, and is thus dominant, in moderate and severe AD. Detailed appraisal again suggests that considerable caution is required in accepting this result with simple sensitivity analyses conducted by the report authors indicating ICERs which would not normally be considered cost-effective by NICE.

1.4.3. PenTAG modelling results

Despite modifications to overcome problems highlighted in the last appraisal, the results of the PenTAG model were not dissimilar to the results for the last TAR, indicating that neither AChEIs nor memantine are cost-effective irrespective of the severity of AD being considered. This is attributable to failing to find cost-savings when the anti-AD treatments are employed, coupled with much smaller modelled estimates of health benefit relative to the manufacturers' submissions. It needs to be highlighted that the incremental effectiveness and cost underlying the ICERs are very small and that the results are highly uncertain and very sensitive to changes in several model assumptions and parameters.

In considering the strengths and weaknesses of the PenTAG model-based analyses, compared with the manufacturer and other models (see below), there should be no initial presumption that the model from the independent review group is somehow more valid or reliable than the others. Rather, in this complex disease area, the diversity of models - and resultant variation in the cost-effectiveness estimates - is partly a reflection of evident structural uncertainty regarding how to simulate this disease and its consequences, as well as differences in the rationales and context for developing each model. The PenTAG model, for example, has been developed in four to five months, with particular expectations to address some of the identified weaknesses of the previous model, and to be a single model capable of evaluating all the treatment comparators at different levels of disease severity (both the AChEIs and memantine). The manufacturers, in contrast, have had a longer time period in which to develop their models, full access to their own trial data with which to inform them, and the more specific goal of evaluating the cost-effectiveness of their product.

The probabilistic sensitivity analyses suggested rivastigmine patches ($10cm^2$) were the most cost-effective of the AChEIs, but only with a probability of 17% at a willingness to pay of £30,000 per QALY (15% at a willingness to pay of £20,000 per QALY). Best supportive care was the most cost-effective option with a probability of 57% of being cost-effective at a willingness to pay of £30,000 per QALY (62% at a willingness to pay of £20,000 per QALY). When compared to the next cheapest, non-dominated technology, the estimated deterministic ICER for rivastigmine patches compared to best supportive care was £61,100 per QALY, and galantamine (16-24mg) was associated with an ICER of £157,800 per QALY compared to rivastigmine patches. Both donepezil (10mg) and rivastigmine capsules (9-12mg) were dominated. These ICERs should be interpreted with caution in light of the very

small incremental costs and benefits and the considerable parameter and structural uncertainty in the PenTAG model.

For the treatment of moderate to severe AD, the probabilistic sensitivity analysis estimated a probability of less than 4% that memantine was the most cost-effective option when compared to best supportive care at a willingness to pay of £30,000 per QALY (probability of 2.6% at a willingness to pay of £20,000 per QALY). The deterministic ICER was estimated to be £248,500 per QALY for a moderate to severe cohort. Although a great deal of parameter and structural uncertainty was also present in the cost-effectiveness analysis of memantine, none of the alternative assumptions assessed lead to a positive net benefit for memantine compared to best supportive care at a willingness to pay of £30,000 per QALY.

Again it must be repeated that although the ICERs of all drugs are large relative to best supportive care, the incremental net benefits per patient are extremely small, given that the incremental costs and benefits are very small. This implies that funding all the drugs would reduce the total net benefit of the health service only by a very small amount.

1.5. Discussion

1.5.1. Strengths and limitations of the systematic review of studies of effectiveness

The strengths of this systematic review are that is was conducted by an independent research team using the latest evidence.

There are a number of limitations:

- The length of follow up of the trials was a maximum of six months, which makes it very difficult to reliably extrapolate findings years ahead.
- There is a lack of evidence from the trials on key outcomes such as mortality, institutionalization, the impact on carer's time and the prescription of anti-psychotics.
- None of the trials conducted sub-group analyses based on disease severity, making us unable to comment on the effectiveness of treatments for mild, moderate or severe AD separately.

- Overall the quality of the trials was moderate to poor, with lack of reporting of key measures of trial quality, thus adding to the uncertainty of the results.
- The use of LOCF and OC methods for accounting for missing data may have overestimated the treatment benefit from the drugs.
- Some of the measures used in the trials are insensitive to change in Alzheimer's disease (ADAS-cog, MMSE). Therefore, the effects of treatment may have been underestimated in some cases.
- The searches were limited to the English language due to resource limitations, which may have led us to exclude important studies.

1.5.2. Strengths and limitations of the economic modelling by PenTAG

Although we believe we have made a number of improvements on the previous SHTAC-AHEAD model, and attempted to address some of the specific criticisms of the previous model (as detailed in Appendix 17), it should still be regarded as a explorative model for assessing the cost-effectiveness of drug treatments in this highly complex disease area. The main reasons for viewing the updated model and its outputs with such caution are:

- The underlying disease model captures just the two dimensions of cognitive status and functional status/ADL. Behavioural and psychological symptoms are not incorporated into the model, and therefore any treatment effects and quality of life impacts related to these symptoms will not be captured.
- The expression of treatment effectiveness, while based on a multivariate formula based on patient age, ADL status and cognitive status, is mainly based on predicting delays in time-to-institutionalization. While there is good evidence that this event/transition marks a key change in care costs, the evidence that it is also a key marker of decline in quality of life is uncertain.
- Although the model now incorporates more graduated declines in patient utility, and more graduated increases in NHS and PSS costs prior to institutionalization, assuming that all of these time-related cost and utility changes will be delayed by the same amount of time that institutionalization is delayed is a key assumption in the model

(especially bearing in mind that many of the health care costs will not be related to Alzheimer's)

- The main database of individual patient data from the UK that the time-to-institutionalization model and key cost parameters are largely based upon is relatively old (1988-1999), small (n=92 with AD) and from a small part of the UK (Oxfordshire). Its generalisability to England and Wales in 2010 therefore has to be considered.

Unlike the 2004 SHTAC analysis, utility benefits pre-institutionalization have been accounted for since utilities are based upon MMSE, and both costs and MMSE prior to institutionalization are conditional on time until institutionalization. However, as with the previous model, basing the simple structure of the model around the two main stages of living in the community (i.e. at home), or living in a nursing or residential home (or long-term hospitalisation), means estimating the benefits of drug treatments for those already in residential care is problematic. This is a more considerable weakness of this modelling approach for evaluating the cost-effectiveness of memantine.

In attempting to overcome a criticism of the SHTAC model where AD progression was based on US data, AD progression in the PenTAG model is based on UK individual patient data. However, the generalisability of this data should be questioned for a number of reasons: (i) the data are from just 92 individuals, (ii) it is collected from Oxfordshire only, and (iii) these data are now rather out of data, as they were collected between 1988/9 and 1999. Not only are these data used to inform AD progression, they are also used as a basis for the NHS/PSS costs of care (in the community and in institutions). This has an advantage in one respect since there is no need to incorporate an additional source of evidence, with its own uncertainties, into the model. However if the data from Wolstenholme and colleagues cannot be generalised to England and Wales in 2010, it is likely the model will not be generalisable either, even though few options were available as the basis for predicting disease progression. In addition to considering the US data used in the SHTAC model to model disease progression, RCT data were considered but felt not to be ideal due to the restricted populations from inclusion/exclusion criteria. The available UK epidemiological evidence was either from Wolstenholme and colleagues or a longitudinal cohort study where many participants were receiving AChEI and/or memantine treatment (i.e. the LASER-AD study).

The incorporation of the full treatment effect at six months is artificial. It is more likely that improvements due to treatment are gradual. It is also assumed in the PenTAG model that

treatment benefits remain after treatment has ceased. This assumption is also likely to be unrealistic, but is favourable to the active treatments. Furthermore, the treatment effects incorporated into the PenTAG model are absolute effects: there has been no accounting for differential effects for baseline severity, but there was some, albeit explorative, evidence of an association between baseline MMSE and functional outcomes identified in *(Appendix 7).*

A further limitation relates to effectiveness data availability. No relevant ADL data for donepezil and no relevant MMSE data for galantamine at 21-26 weeks were identified from the clinical effectiveness review. It was assumed that this was a lack of evidence for an effect, rather than lack of effect and a class effect was assumed (i.e. the effectiveness was assumed to be the same as the other AChEIs).

1.6. Conclusions

The additional clinical effectiveness evidence identified in this up-date systematic review continues to suggest that there is clinical benefit from the AChEI's in alleviating symptoms and controlling disease progression in Alzheimer's disease. However, there is only randomised evidence for this up to six months. Although there is also new evidence on the effectiveness of memantine, it remains less supportive of this drug's use.

While there remains considerable debate about the magnitude of the effect of AChEIs on cognition, function, behaviour and global impact, there is very little, if any, disagreement that the effects are present.

Conclusions concerning cost-effectiveness are however no clearer. This arises from uncertainty about the most appropriate modelling approach, compounded by uncertainty about all model parameters. Although we can explain some of the large differences in the cost-effectiveness estimates between the industry submissions for donepezil and memantine and the PenTAG model, these cannot be completely accounted for.

Whatever the final judgement about the most likely true ICER values, it must be recognised that the estimates are based on very small incremental benefits and costs.

1.6.1. Implications for service provision

These are not clear and will ultimately rest on the interpretation of the new evidence from a variety of sources, including this report, in the forthcoming NICE appraisal on this topic.

1.6.2. Suggested research priorities

New research in the following areas could reduce the uncertainty noted:

- Good quality longer term RCTs (following CONSORT) to include mortality, time to institutionalization and HR QOL as outcomes and sufficiently powered for subgroup analysis by disease severity, response to treatment, behavioural disturbance and comorbidities. We have identified that a limited number of major RCTs addressing relevant issues such as management when patients fail to respond to AChEs are already in progress (DOMINO-AD).

- Such good quality trials should aim to use the same standardised measures of cognitive status, functional status/ADL, and behavioural/psychiatric symptoms.
- Systematic reviews of non-RCT evidence on the impact of anti-AD treatments on resource use, institutionalisation and mortality.
- Further independent comparison of different methodological approaches to modelling the cost-effectiveness of anti-AD treatments.
- Research into cognitive measures that are sensitive to change in dementia.
- Studies should measure HRQoL with the DEMQOL which has been validated for use with dementia patients rather than the EQ-5D which has not. Work is needed to derive utility values from the DEMQOL or to map it onto the EQ-5D or HUI 2/3

In addition this report highlights some wider methodological issues which would benefit from further investigation:

- Research into more valid ways of accounting for missing data than LOCF and OC particularly in degenerative diseases like AD.

2. Background

2.1. Aim of the review

The aim of this assessment is to review and update as necessary, NICE guidance to the NHS in England and Wales on the clinical and cost-effectiveness of donepezil, galantamine, rivastigmine, for mild to moderate Alzheimer's disease, and memantine, for moderate to severe Alzheimer's disease, which was issued in November 2006 and amended in September 2007 and August 2009.[1]

This previous guidance was primarily based on evidence presented to NICE in the assessment report by Loveman and colleagues in 2004.[2] We will summarise the evidence presented in this previous report and review and report new evidence from 2004 to the present.

2.2. Description of health problem

2.2.1. Pathology

Definitions

Dementia is usually a disease of later life and has been defined as:

> "a syndrome consisting of progressive impairment in memory and at least one other cognitive deficit (aphasia, apraxia, agnosia, or disturbance in executive function) in the absence of another explanatory central nervous system disorder, depression, or delirium." DSM-IV Diagnostic and Statistical Manual of Mental Disorder, 4th edition.

People with dementia may also show other symptoms such as depression, psychosis, wandering and aggression.

Alzheimer's disease is the most common form of dementia, and is additionally characterised by the presence of neurofibrillary tangles and amyloid plaques in the cerebral cortex, observed at post-mortem.

Diagnosis

In the distant past, diagnosis of Alzheimer's disease before death had been on the basis of excluding other causes ruling other causes out. However, there are now agreed criteria that accurately predict up to 90% of Alzheimer's disease cases (see *Table 1*). Alternatively, diagnosis can be made from ICD-10* and DSM-IV†.

TABLE 1 *The NINCDS-ADRDA* criteria for Alzheimer's disease*

Probable Alzheimer's disease
▪ Dementia established by clinical examination, documented by the Mini-Mental State Examination or similar and confirmed by neuropsychological tests
▪ Decline in memory and at least one non-memory intellectual function
▪ Decline from previous level and continuing decline
▪ Onset between 40 and 90 years of age
▪ No disturbance in consciousness
▪ Absence of systemic disorders or other brain diseases that in and of themselves could account for the progressive deficits in memory and cognition
Definite Alzheimer's disease
▪ Clinical criteria of probable Alzheimer's disease
▪ Histopathological evidence of Alzheimer's disease at post-mortem or biopsy
Possible Alzheimer's disease
▪ Patient has dementia syndrome with no other cause but clinical variation from typical Alzheimer's disease
▪ Patient had second disorder that is sufficient to produce dementia but not considered the cause of the dementia
▪ Single gradually progressive cognitive deficit in absence of other causes

* NINCDS-ADRDA, National Institute of Neurological and Communicative Disorders and Stroke-Alzheimer's disease and Related Disorders Association.3

Source: Warrell DA, Cox TM, Firth JD Eds. Oxford Textbook of Medicine. Fourth Edition. 2003

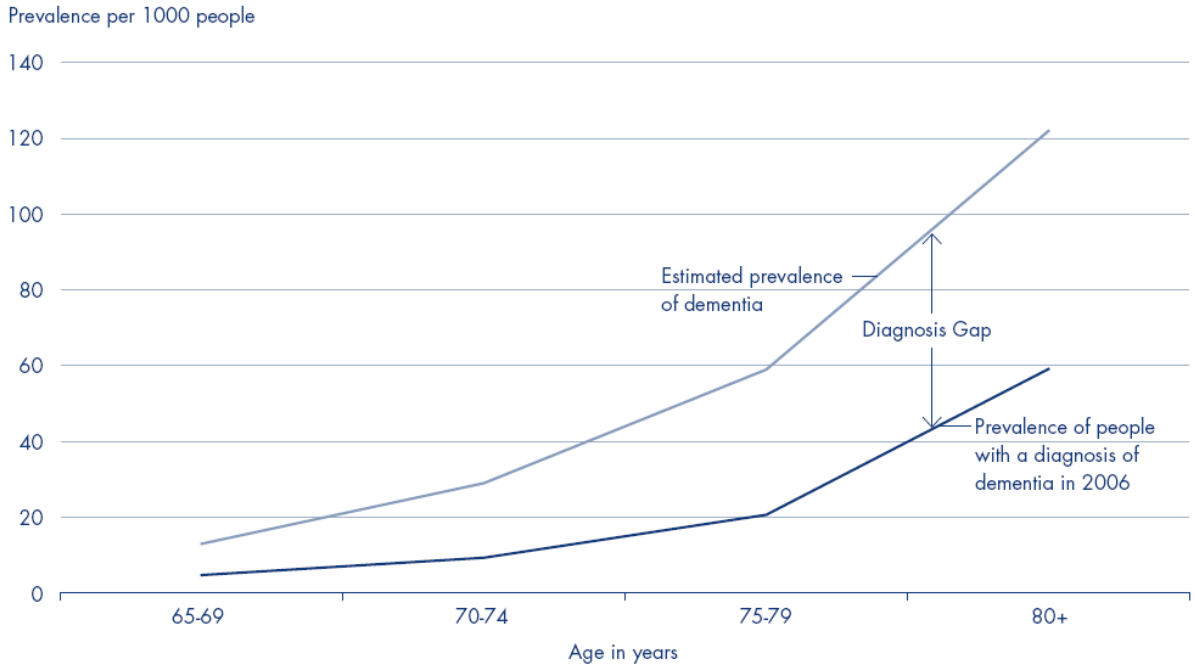
The diagnosis of dementia may happen many months after onset as the development of symptoms and is usually insidious. It may take some time for the individual to realise that significant memory, mood or ability changes are taking place. Other possible diagnoses, such as depression, delirium, vitamin B12 deficiency, hypothyroidism have to be excluded first. Further testing is necessary to determine the particular cause of dementia.[4]

* DSM-IV, Diagnostic and Statistical Manual of Mental Disorders, fourth edition

† ICD-10, International Classification of Diseases, 10th revision

Examination of data from Primary Care Trusts, the Dementia UK report (2007) and the Office of National Statistics by the National Audit Office has indicated that, in England, more than 50% of people with dementia never receive a correct diagnosis.[5] See *Figure 1* for 2006 estimates of the diagnosis gap.

FIGURE 1 *The gap between prevalence and diagnosis of dementia in England*



Source: Knapp et al. (2007) Dementia UK: Report to the Alzheimer's Society, King's College London and London School of Economics and Political Science (estimated actual average prevalence) and General Practice Research Database report to the NAO (reported prevalence based on diagnoses)

NOTE

The graph above shows reported prevalence of dementia, based on levels of diagnosis within Primary Care Trusts, for ages 65 and upwards in 2006. The estimated actual average prevalence has been calculated using data from 2007 *Dementia UK* report in conjunction with population estimates from the Office for National Statistics. The latter does not take into consideration those aged 85 and above, due to restrictions on the data available, and as such forms a very prudent estimate of the true prevalence of dementia in the over 65 population.

Source: Knapp et al. 2007 in, National Audit Office: Improving services and support for people with dementia. 2007

2.2.2. Epidemiology

Alzheimer's disease is predominantly a disease of later life with some 5% of the UK population over 65 years affected. Early onset Alzheimer's disease can be found in younger people, this is a rare condition, only accounting for an estimated 2.2% of those with dementia.[6] Currently, there are estimated to be about 820,000 people in the UK with dementia (1.3% of the population), and of these approximately 520,000 (62%) will have Alzheimer's disease, of these, approximately 423,000 (83%) live in England and 26,000 (5%)

live in Wales.[6;7] Alzheimer's disease is more commonly found in women than men in the UK, with 67% of women with dementia having Alzheimer's disease but only 55% of men.[6] However, the association with gender is completely explained by the shorter life-expectancy of men.[8]

The incidence of dementia therefore increases with age. In England and Wales, for people aged 65 to 69 years, the incidence is estimated to be 7.4 (95% CI 3.6-16.1) per 1,000 person years, this rises to 84.9 (95% CI 63.0-107.8) per 1,000 person years at 85 years old and above.[9] These rates predict 180,000 new cases of dementia per year, and if 62% of these have Alzheimer's disease (see above) then there are approximately 111,600 new cases of Alzheimer's disease in England and Wales per year. The Medical Research Council's Cognitive Function and Aging Study (2006) found, that in England and Wales, increasing age was the greatest risk factor for dementia, with gender weakly associated. Having Parkinson's disease increased the risk of dementia by three times, odds ratio 3.5 (95% CI 1.3-9.3), but rating your own health as poor was a greater risk factor, odds ratio 3.9 (95% CI 2.2-6.9). Better education was a marginally protective factors, 0.7 (95% CI 0.5-1.0).[8]

Table 2 shows the combined numbers of diagnosed and undiagnosed cases of dementia in the UK in 2006 estimated by the Dementia 2010 study.[7]

TABLE 2 *Number of diagnosed and undiagnosed dementia cases in the UK in 2006*

Age group, years	Male	Female	Total
30-59	19,840	11,381	31,221
60-64	25,034	7,782	32,816
65-69	28,056	15,378	43,434
70-74	50,085	48,319	98,404
75-79	42,805	74,037	116,842
80-84	68,343	120,482	188,825
85-89	50,439	124,465	174,903
90-94	28,399	78,606	107,006
95-99	5,008	23,424	28,432
Total	318,010	503,874	821,884

Source: Dementia 2010: The economic burden of dementia and associated research funding in the UK

2.2.3. Aetiology

The cause of Alzheimer's disease is uncertain. However, it is generally believed that the condition develops from multiple factors, with increasing age bringing the greatest risk. Up to 5% of cases are linked to genetic causes; medical history and lifestyle are also contributing factors.[10] At least three genes have been identified that are associated with the rare condition of early-onset Alzheimer's disease.[11-13] A genetic link is also likely for those with a family history of late-onset Alzheimer's disease, although, a particular gene for this has not been identified.[4]

There is evidence that it may be possible to prevent some incidence of Alzheimer's disease; it is thought that due to the cerebrovascular contribution to brain pathology, that managing cardiovascular risk factors (high cholesterol, high blood pressure, type II diabetes and being overweight) may delay or prevent the onset of Alzheimer's disease. Other possibly preventative factors include, regular exercise, a low fat diet and a good social network.[14-16]

2.2.4. Prognosis

There is currently no cure for Alzheimer's disease. There is variation in the time it takes from diagnosis to death. The estimated median survival for Alzheimer's disease from onset has been calculated as 7.1 years (95% CI 6.7-7.5 years) in the USA by Fitzpatirick and colleagues[17] and is reported in Warrell and colleagues as about 10 years in the UK.[18] Although, survival figures are varied and depend on whether they are from time of reported onset to time of actual diagnosis, in general a diagnosis of Alzheimer's disease halves life-expectancy.

The contribution of Alzheimer's disease to these survival figures is difficult to know, as people with Alzheimer's disease frequently have co-morbidities which will influence their longevity. The proportion of deaths estimated to be due to Alzheimer's disease increases with age and varies with gender. At 65 years old 1% of women and 2% of men are likely to die from dementia, at 85-89 years old this rises to 23% of women and 18% of men.[6]

2.2.5. Impact of health problem

2.2.5.1. Significance for patients

It can take several years of slow deterioration for the full effects of Alzheimer's disease to be felt.[19] In the early stages there can be severe memory loss for recent events with associated repetitive questioning and loss of the ability to learn.[19;20] There may be a general deterioration in the ability to socialise which can be difficult for both sufferer and carer to cope with.[21;22] As mild Alzheimer's disease takes hold, normal activities of daily living, such as shopping or managing finances, become increasingly difficult as cognitive function deteriorates.[23] Communication also becomes a problem as vocabulary shrinks and fluency falters.[23;24] At this stage the sufferer may still be aware of their failing abilities and the experience and known outcome of Alzheimer's disease can frequently lead to associated depression.

Disease progression to moderate Alzheimer's disease leads to further loss of cognitive abilities, including the ability to remember and/or understand words. Activities of daily living become increasingly affected as the ability to perform purposeful movements decreases e.g. getting dressed or cooking. Commonly there are also neuropsychiatric symptoms such as anxiety, wandering, irritability, disinhibition and apathy. Visual and auditory hallucinations

occur in about 30% to 59% of sufferers.[18] Managing these symptoms can be a very difficult burden for carers, who may well be elderly themselves. Indeed, the main predictors of full time institutional care are caregiver exhaustion,[25] the degree of patient dependence[26] and the rate of disease progression.[27]

In developed countries sufferers of Alzheimer's disease usually end their days in institutional care as the last stages of Alzheimer's disease bring complete dependence. This final stage is characterised by limitations such as: inability to walk; manage personal care; mutism; inability to recognise familiar people and objects and incontinence. There may also be seizures and involuntary twitching.

2.2.5.2. Significance for carers

Being the main carer for a person with Alzheimer's disease can have an enormous impact on physical, psychological and social well being.[28] From the early frustrations, prior to diagnosis, of living with others' impaired cognitive function, through the devastating diagnosis, to the knowledge that the relation/friend is going to get progressively worse and die, the outlook for carers is bleak. Many carers are elderly spouses, perhaps with health concerns of their own, or grown-up children who now have their own families to care for as well.[29] Carers may cope reasonably well with the early stages of the disease but as the behavioural and psychological symptoms of dementia (BPSD) become more severe, full-time institutional care becomes increasingly likely.[4] For some this brings feelings of guilt and depression,[30;31] possibly leading to the cognitive decline of the carers themselves.[32] Behavioural and psychological symptoms are common in Alzheimer's disease and may be difficult to manage, causing distress to carers and patients alike. They have been shown to be better predictors of institutionalisation[33] and carer distress[34] than cognitive symptoms.

As Alzheimer's disease progresses increasing grief and feelings of loss may be experienced by carers.[35] Findings from the Eurocare European study of co-resident spouse carers of dementia sufferers showed that co-resident carers carried a heavy burden and that mental distress was high. They concluded that issues of behavioural disturbance, negative social reactions, financial worries and younger spouse carers predicted greater distress.[36] However, there is evidence that enhanced counselling and support can relieve symptoms of depression in caregivers and delay admission to institutional care of the patient.[37]

2.2.5.3. Significance for the NHS and Social Services

With an increasingly elderly population the burden of Alzheimer's disease upon the NHS and Social Services is considerable. Of the estimated 520,000 people in England and Wales with Alzheimer's disease,[7] It is estimated that approximately 63.5% live at home and 36.5% are in residential care.[6] Unsurprisingly the risk of moving into residential care increases with age and disease severity. The proportion of people with severe dementia increases from 6.3% for those between 65-69 years old to 23.5% for those aged 95 years or older.[6] Consequently, the proportion of people in the UK with dementia who live in residential care rises from 26.6% of those aged 65-74 and to 27.8% of those aged 75-84 years, to 40.9% of those aged 85-89 years and 60.8% of those aged 90 years or older.[6]

As the disease progresses the balance of burden of care shifts from predominantly falling on the informal carers, to the NHS and Social Services as patients are sustained with medication and support at home, until finally financial costs fall mostly on Social Services as patients move into institutional care, although a proportion of this cost may be borne by the carer or their family. A longitudinal cohort study by Banerjee and colleagues (2003) has found that when a person with dementia lives with their main carer they are 20 times less likely, over the course of a year, to move into residential care than those who do not, odds ratio 0.05 (95% CI 0.01 to 0.42). They also found the carer's psychological quality of life and the severity of behavioural problems shown by the patient were predictors of institutionalisation, odds ratios 1.10 (95% CI 1.02 to 1.19) and 1.08 (95% CI 1.01 to 1.15) respectively.[38] Although de Vugt and colleagues found, in a similar study, that it was the carer's response to the behavioural symptoms, rather than the symptoms themselves that predicted institutionalisation.[39]

2.2.5.4. Measurement of disease

Details of individual measures used in the included trials can be found in Appendix 1.

A review of outcome measures used in clinical trials of drugs for Alzheimer's disease by Wolfson and colleagues (2000)[40] revealed a number of shortcomings in these measures. In particular they found that several of the scales had weak psychometric properties e.g. lack of responsiveness to change. Some studies had small sample sizes and others used inappropriate statistical analyses.[41]

The progress and symptoms of Alzheimer's disease can be measured through cognitive tests, behavioural measures, measures of functional ability/quality of life, and global rating scales.

A thorough assessment of cognitive ability would include measures of attention, processing speed, visuospatial function, praxis, language, executive function and abstraction. The most commonly used scales for this domain are the Mini-Mental State Examination[42] (MMSE) and the Alzheimer's disease Assessment Scale-cognitive[43] (ADAS-cog). While the MMSE's validity and reliability as a screening tool for Alzheimer's disease have been established,[42] it has problems with identifying change over time and scores are affected by people's level of education.[44;45] Similarly the ADAS-cog has been criticized for its insensitivity to change in cognitive ability at either end of the severity continuum.[46] It is concerning that the most commonly used instruments to measure change in cognitive function in drug trials for Alzheimer's disease should be insensitive to change.

The measurement of behaviour change is important as it is these symptoms that many care givers find most difficult to cope with, precipitating the transition into institutional care.[33;34] The most frequently used measure of behavioural change in Alzheimer's disease trials is the Neuro-psychiatric inventory (NPI).[47] This is a proxy-rated scale usually completed by the main carer; its validity and reliability have been demonstrated by Cummings and colleagues.[47]

There are two kinds of global rating scales for Alzheimer's disease; those that measure the severity of illness at a point in time e.g. the Clinical Dementia Rating scale (CDR)[48] and can, if used repeatedly, plot mental deterioration over time. The other sort of global instruments are change scales, such as the Clinician Interview-Based Impression of Change plus Caregiver Input (CIBIC-plus).[49] These measure broad changes in Alzheimer's disease. However, their use may be biased towards cognitive abilities, as Claus and colleagues have found that clinicians may have a bias towards this aspect of Alzheimer's disease, whilst carers place more emphasis on behavioural and psychological symptoms and functional ability.{Claus, 1998 10183 /id} However, the use of CIBIC-plus may help to overcome this.

Measures of functional status in clinical trials are most commonly taken using the Activities of Daily Living scale[51] (ADL) or the Instrumental Activities of Daily Living (IADL).[52] Their reliability and validity has been described by McDowell and Newell.[53] However, this is not in the specific context of dementia.

Although the DEMQOL has been validated as a measure of health related quality of life in people with dementia,[54] in clinical trials the most frequently used measure is the Patient-rated QoL scale. This is a seven item patient-rated scale that measures feelings of well-being in the domains of, relationships, eating, sleeping and social and leisure activities, on a 0-50 analogue scale.[55]

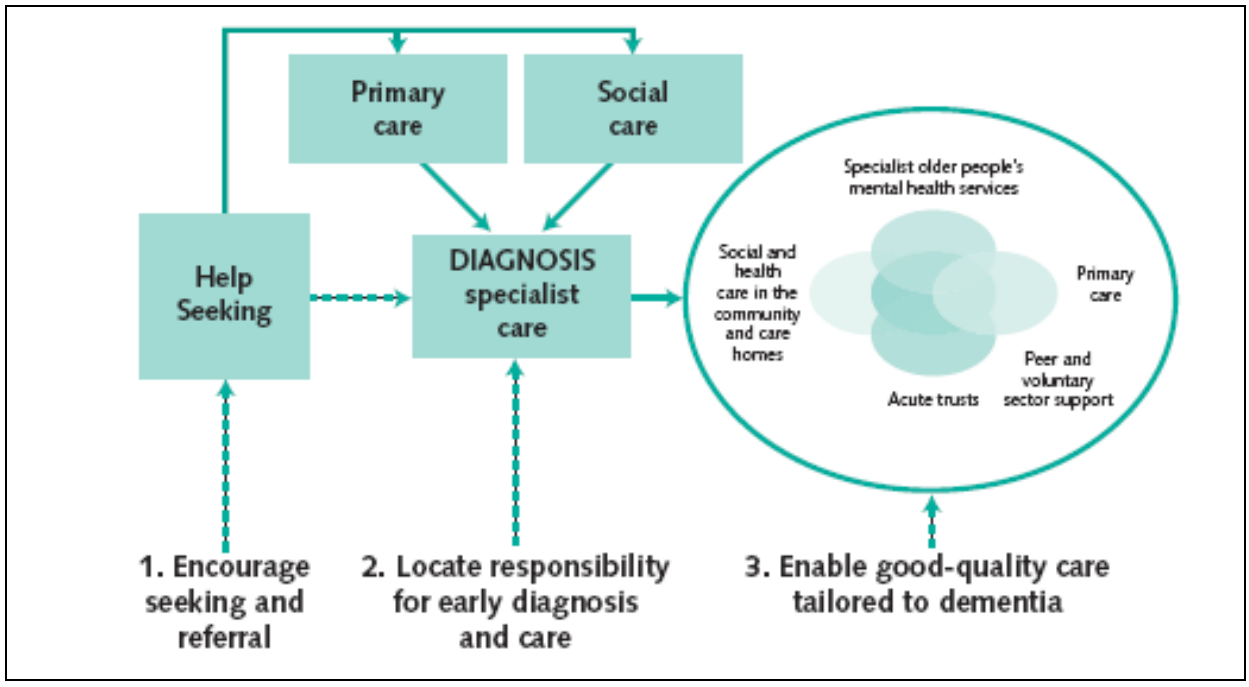
2.2.6. Care for people with Alzheimer's disease

The National Dementia Strategy (2009) says that everyone with suspected dementia should have access to:

A rapid and competent specialist assessment; an accurate diagnosis sensitively communicated to the person with dementia and their carers; and treatment, care and support provided as needed following diagnosis. The system needs to have the capacity to see all new cases of dementia in the area.[56]

In order to achieve this goal the Department of Health has set out the following care pathway.[56] See *Figure 2*.

FIGURE 2 *Care pathway summarising the three themes of the National Dementia Strategy and the commissioning challenges*



Source: Living well with dementia: A National Dementia Strategy. 2009.

The provision of care for people with Alzheimer's disease is complex as it is shared between informal voluntary care, private care, Social Services and the NHS.

2.2.6.1. Informal care

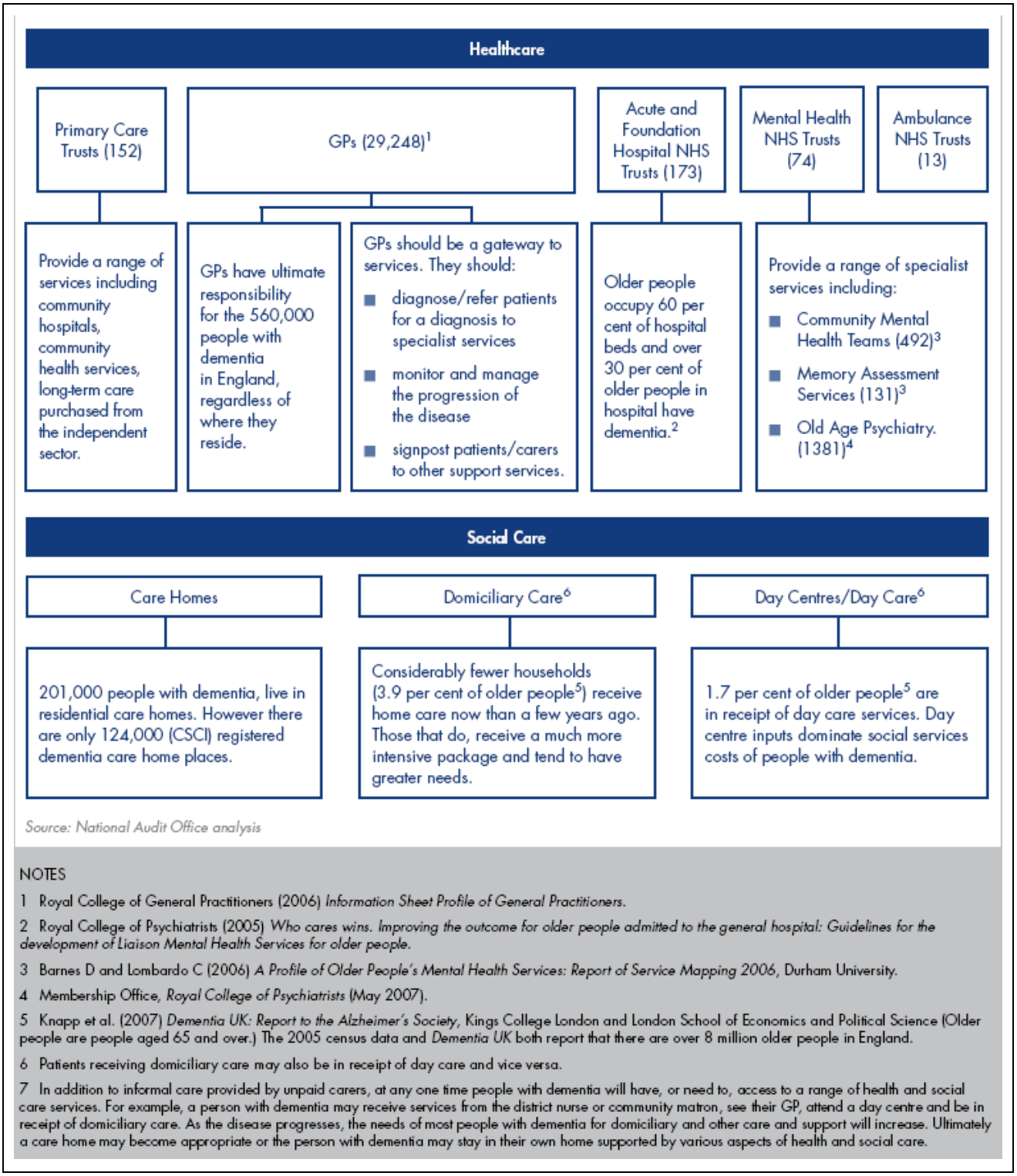
Carers

An analysis of General Household Survey (1998/99) data estimated that 53% of people over 65 who could not live completely independently were supported by unpaid carers.[57] This estimate translates to approximately four million carers in England, most of working age.[58] Changing demographic patterns, with children living a considerable distance from their parents and more single people, may mean that this caring resource is reduced in future.[59] Reports in the last decade have promoted support for carers; Support for Carers of Older People;[58]Caring about Carers: A Strategy for Carers in Wales (Implementation Plan),[60] and, The NHS Plan.[61] However, many carers feel unsupported and isolated.[58] The burden of caring can affect the health and well-being of carers,[62] possibly with high levels of depression,[63] although another study found that over a two year period carer's psychological well-being did not deteriorate.[64] Another effect of caring is the reduction of the capacity of the carer to earn a living.[65] The Medical Research Council's Cognitive Function and Ageing Study (MRC CFAS) also found that 9% of carers of people with dementia had reduced their hours of work and one fifth of carers who were younger than the statutory retirement age had given up work completely.[64]

2.2.6.2. Formal care

The formal care of people with Alzheimer's disease falls mainly to the NHS and Social Services, although private and voluntary sector agencies are also involved. The National Audit Office has produced the following diagram to show the types of providers that are currently involved in dementia care. See *Figure 3*.

FIGURE 3 *Key public providers involved in the formal care of people with dementia.*



Source: Improving services and support for people with dementia (2007). National Audit Office

NHS Services

The medical needs of people with Alzheimer's disease span a wide range of specialities; as people with Alzheimer's disease are usually elderly, co-morbidities are common and their treatment is frequently complicated by the dementia. However, the main specialty involved is old age psychiatry, although, in the UK, a geriatrician may be responsible for diagnosis and treatment. The first contact a patient has is usually with primary care; a correct and early provisional diagnosis here is vital as this is the way into specialist care.[6]

There are no nationally agreed criteria for referral; therefore the burden of care falls differentially on primary and secondary care depending on location. The kinds of care provided can be grouped as either those for 'serious mental illness' or 'early intervention', depending on the severity of symptoms. The initial assessment of someone who may have Alzheimer's disease is ideally conducted in their home, although many people with early stages of the disease are now seen in memory clinics. The home is a preferable setting to an out-patient department because it enables the assessor to see how the person functions in everyday situations. It also enables risk assessment of potential dangers in the home and is more likely to take place, as the possibly confused and forgetful patient may lack understanding of their need to attend an assessment appointment.

Social Services

Apart from cognitive and psychological decline, people with Alzheimer's disease face a gradual loss of their ability to live independently. Initial support with everyday activities frequently comes from family and friends. However, where this is not available, and when the disease progresses, such support predominantly comes from Social Services, although private agencies may be involved.

There are no statistics about the total number of people with Alzheimer's disease who are supported at home either by Social Services or the private sector. However, in England, there has been an increase in recent years, in the volume of home-care bought by local authorities, a decrease in the numbers of people supported and an increase of support from private providers.[6] This means that fewer people are receiving help at home from Social Services than in the recent past, but those who do generally have greater needs and are receiving more comprehensive support. A consequence of this is that people are entering

full-time residential care at later stages of Alzheimer's disease. In Wales the picture is different, with a decrease in the amount spent on home-care by local authorities.[6]

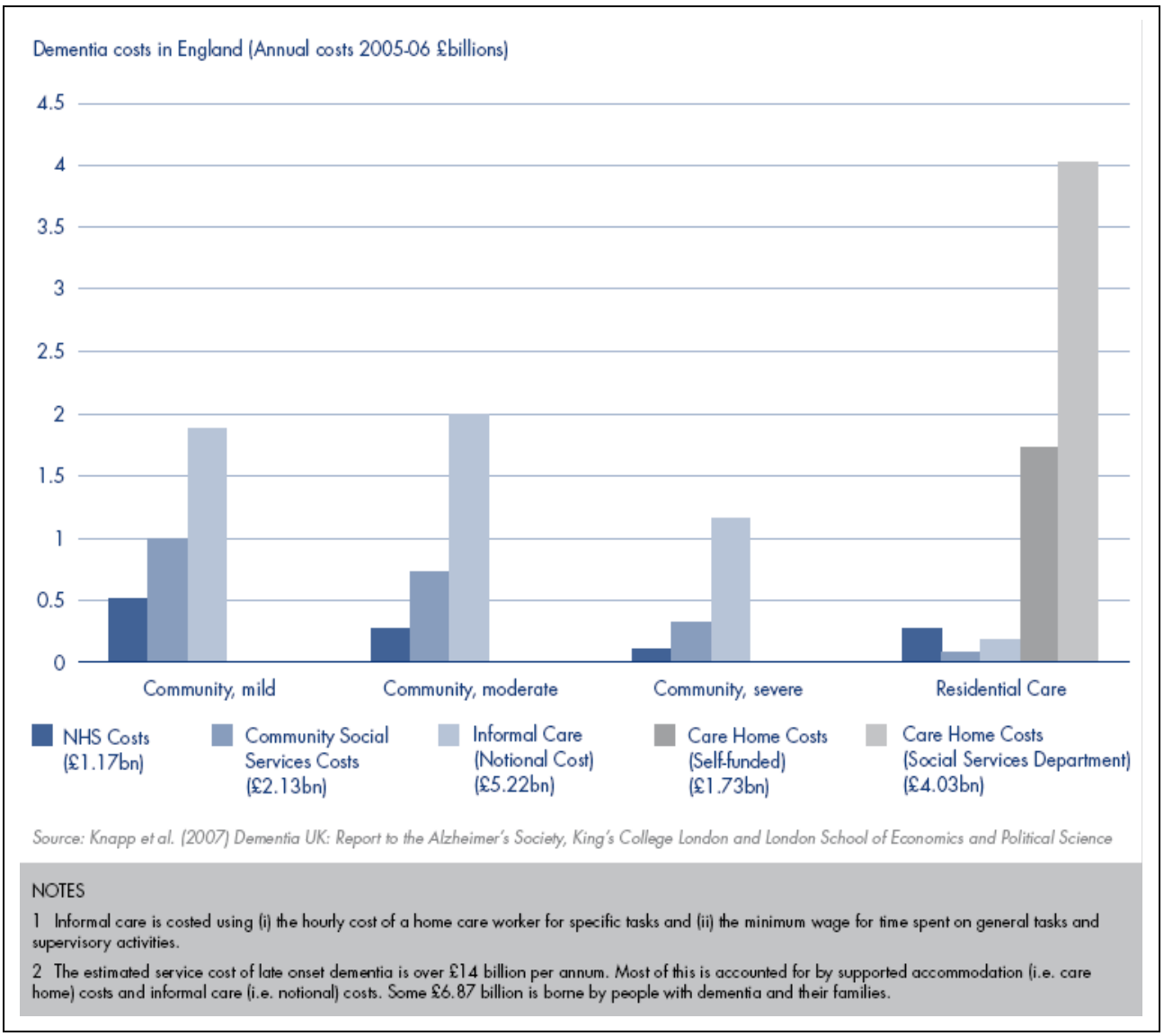
In recent years the supply of residential care homes has been in decline in England and the balance of ownership has changed with more homes now being in private hands.[6] Also the average size of care home has increased to 34 beds but the quality of care provided continues to be variable; areas of concern include unstimulating environments and a low paid, poorly trained work-force that has a high turn-over, which undermines the building of relationships between staff and residents.[58] Standards have begun to rise due to regulation, according to the Audit Commission, but there is a long way to go.

2.2.6.3. The cost of care- overview

Around two thirds of the care for people with Alzheimer's disease comes informally from the community and it is the family who bare the greatest burden of cost.[5] The following diagram from the National Audit Office report (*Figure 4*) shows how the cost of dementia was spread in England in 2005/6.*

* The categories of mild, moderate and severe dementia are based on the Cambridge examination for mental disorders of the elderly (CAMDEX).[66]

FIGURE 4 *Dementia costs in England 2005/6 by severity and place of residence*



Source: National Audit Office: Improving services and support for people with dementia. 2007

It has been estimated in the recent Dementia 2010 report that these costs have risen to an annual cost to the UK economy of £23 billion per year (2007/08), (this compares with £12 billion per year for cancer and £8 billion per year for heart disease). The majority of this £12.4 billion cost fell on unpaid carers, (55%), Social Services funded 40% (£9 billion) of the cost and the NHS 5% (£1.2 billion).[7]

Presuming that 62% of people with dementia have Alzheimer's disease – see Section *2.2.2* above, this translates to an annual cost to the UK economy for Alzheimer's disease of over £14 billion per year. For each Alzheimer's patient this gives an estimated annual cost of

£27,647, more than cancer (£6,000), stroke (£5,000) and heart disease patients (£3,500) put together.[7] A full description of the costs of care for people with Alzheimer's disease can be found in Section 8.3.10.2.

2.2.6.4. Variation in services

The Dementia UK report (2007) indicated that there is a wide variation in the prescribing of anti-dementia medication in England and Wales.[6] Information specific to Alzheimer's disease was not available, but it may be reasonable to suggest that the picture would be similar. The data were collected by IMS, a medical information consultancy, between October 2005 and September 2006 from 50% of the pharmacies in England and Wales and represent 90% of all UK prescribing. This information is a reflection of national commissioning practice and shows, by PCT, the likelihood of being prescribed medication for dementia. The number of prescriptions per person with dementia in primary care varied from 12.0 prescriptions per year in Knowsley to 0.4 in West Berkshire. Most PCTs (75%) prescribed between 1.0 and 4.0 prescriptions per year.[6] The reason for this variation in provision is unclear.

2.2.7. National guidelines, guidance and reports

The following national guidelines, guidance and reports are related to this technology appraisal.

- Dementia 2010: the economic burden of dementia and associated research funding in the United Kingdom[7]
- Living well with Dementia: a national dementia strategy (2009)[56]
- NICE technology appraisal guidance 111: Donepezil, Rivastigmine, Galantamine (review) and Memantine for the treatment of Alzheimer's disease (amended August 2009)[1]
- Dementia UK: the full report (2007)[6]
- Dementia: The NICE-SCIE Guideline on Supporting People with Dementia and their Carers in Health and Social Care (2007)[4]
- Improving services and support for people with dementia (2007)[5]

- Everybody's Business- Integrated mental health services for older adults: a service development guide (2005)[67]
- Forget Me Not (2002) – the Audit Commission[68]
- National Service Framework for Older People (2001)[69]
- Forget Me Not 2000 – the Audit Commission[70]

2.3. Description of technology under assessment

2.3.1. Summary of intervention

2.3.1.1. Three licensed AChEIs

This technology assessment report (TAR) will consider four pharmaceutical interventions. Three have marketing authorisations in the UK for the treatment of adults with mild to moderately severe Alzheimer's disease (measured by the MMSE 26-10). These are donepezil (Aricept®, manufactured by Eisai), rivastigmine (Exelon®, manufactured by Novartis), and galantamine (Reminyl®, manufactured by Shire Pharma). They are acetylcholinesterase (AChE) inhibitors, which work by restricting the cholinesterase enzyme from breaking down acetylcholine thus increasing the concentration and duration of acetylcholine at sites of neurotransmission.

Donepezil hydrochloride (Aricept®) is manufactured by Eisai Ltd and co-marketed with Pfizer, it was the first drug to be licensed in the UK specifically for Alzheimer's disease. Donepezil is a reversible, specific, AChEI. Donepezil is easily absorbed by the body and can be taken once a day, initially at 5mg and then, after four weeks use, titrated up to 10mg per day if necessary.

Possible side effects associated with Donepezil include, bradycardia (particularly in people with sick sinus syndrome or other supraventricular cardiac conduction conditions), seizures, nausea, vomiting, diarrhoea, muscle cramp, urinary incontinence, fatigue, insomnia and dizziness.

Rivastigmine tartrate (Exelon®) made by Novartis pharmaceuticals, is a selective inhibitor of acetylcholinesterase and also butrylcholinesterase, another enzyme. Due to its short half-life (1.5hrs) it has to be taken twice a day. Doses start at 3mg per day and increase gradually to between 6mg and 12mg per day. It can be taken orally or by a transdermal patch, with doses of either 4.6mg/24hr or 9.5mg/24hr.

Care should be used with people with renal disease, mild or moderate liver disease, sick sinus syndrome, conduction abnormalities, gastric or duodenal ulcers and a history of asthma or obstructive pulmonary disease. The main possible side effects found are nausea and vomiting, usually in the dose escalation phase.

Galantamine (Reminyl®), manufactured by the Shire Pharmaceuticals Group. Galantamine was originally made from snowdrop and narcissus bulbs but is now synthetically produced. It is a reversible inhibitor of acetylcholinesterase, with a half-life of about seven hours, indicating that it should be taken twice daily at the recommended dose of 16-24mg each time. An alternative version (Reminyl XL) is taken once daily at doses of 8, 16 or 24mg.

The side effects from galantamine are similar to those of the other AChEs and are mainly gastrointestinal; abdominal pain, diarrhoea, nausea and vomiting, although bradycardia and dizziness have been reported.

2.3.1.2. Memantine

The fourth drug, **memantine hydrochloride** (Ebixa®) manufactured by Lundbeck, has a UK marketing authorisation for the treatment of people with moderate to severe Alzheimer's disease (measured by the MMSE, score of 20 or less). It is a voltage-dependent, moderate-affinity, uncompetitive N-methyl-D-aspartate (NMDA) receptor antagonist that blocks the effects of pathologically elevated tonic levels of glutamate that may lead to neuronal dysfunction. Memantine is taken twice a day by mouth. The starting dose is 10mg/day and this can be increased to a maximum daily dose of 20mg/day.

Caution should be used when prescribing memantine for people with renal failure or epilepsy; it is also contra-indicated for people with severe renal impairment. Side effects may include dizziness, confusion, headache, and incontinence.

3. Definition of the decision problem

3.1. Decision problem

The inclusion criteria for this assessment are as follows.

3.1.1. Population

The population for this assessment is adults with Alzheimer's disease. However, as in the assessment which informed TA111, where trials have included participants with mixed dementias, these trials will be included where the dominant dementia is Alzheimer's disease. Papers will be considered on a case by case basis.

3.1.2. Intervention

The intervention to be included is dependent on the severity of Alzheimer's disease, measured by the Mini Mental State Examination criteria:

- mild AD (MMSE 21-26): donepezil; galantamine and rivastigmine
- moderate AD (MMSE 10-20: donepezil; galantamine, rivastigmine and memantine
- severe AD (MMSE <10): memantine

3.1.3. Comparators

The comparators are again dependent on the severity of the Alzheimer's disease.

- mild AD (MMSE 21-26): placebo or best supportive care[a]

[a] Best supportive care: Social support and assistance with day-to-day activities. These include: information and education; carer support groups; community dementia teams; home nursing and personal care; community services such as meals on wheels; befriending services; day centres, respite and care homes.

- moderate AD (MMSE 10-20: donepezil; galantamine, rivastigmine, memantine, placebo or best supportive care
- severe AD (MMSE <10): placebo or best supportive care

3.1.4. Outcomes

The outcomes of interest include measures of:

- Severity of disease and response to treatment
- Behavioural symptoms
- Mortality
- Ability to remain independent
- Likelihood of admission to residential/nursing care
- Health related quality of life of patients and carers (where data permit, analysis will be carried out separately for patients alone, and for patients and carers combined)
- Adverse effects of treatment
- Cost-effectiveness and costs (review of economic studies)

3.1.5. Key issues

All medicines will only be considered according to their UK marketing authorisation.

3.2. Overall aims and objectives of assessment

The purpose of this assessment is to review and update as necessary guidance to the NHS in England and Wales on the clinical and cost-effectiveness of donepezil, galantamine, rivastigmine and memantine, within their UK licensed indications, for the treatment of Alzheimer's disease which was issued in November 2006 and amended in September 2001 and August 2009.

4. Assessment of clinical effectiveness

The purpose of the systematic review of clinical effectiveness is to record the studies found by Loveman and colleagues in 2004[2] and to update their findings with the results of subsequent trials.

This chapter has been arranged as follows:

1. Methods for reviewing effectiveness
2. Results of the systematic review
3. Results: pairwise comparisons
   - **(i)** Donepezil v. placebo
   - **(ii)** Galantamine v. placebo
   - **(iii)** Rivastigmine v. placebo
   - **(iv)** Memantine v. placebo
4. Head-to-head comparisons
5. Combination therapy
6. Results: Multiple Treatment Comparisons
   - **(i)** Cognitive
   - **(ii)** Functional
   - **(iii)** Behavioural
   - **(iv)** Global
7. Summary of clinical effectiveness

4.1. Methods for reviewing effectiveness

The clinical effectiveness of donepezil, galantamine, rivastigmine and memantine for AD was assessed by a systematic review of research evidence. The review was undertaken

following the principles published by the NHS Centre for Reviews and Dissemination (CRD).[71] The study protocol can be viewed on the NICE website, www.nice.org.uk.

4.1.1. Identification of studies

Electronic databases were searched for systematic reviews and/or meta-analyses, randomised controlled trials (RCTs) and ongoing research in November 2009 and updated in March 2010, this updated search revealed no new includable studies. Appendix 2 shows the databases searched and the strategies in full. These included: The Cochrane Library (2009 Issue 4, CDSR and CENTRAL), MEDLINE, MEDLINE In Process, EMBASE, PsycINFO, EconLIT, ISI Web of Science Databases: Science Citation Index, Conference Proceedings Citation Index- and Biosis, the CRD databases: NHSEED, HTA, and DARE databases. Where possible a controlled trials and human filter was added. As this is an update of a previous review the searches were run in the timeframe 2004 to current. The meta-register of controlled trials and clincaltrials.gov were searched for ongoing trials. Bibliographies of included studies were searched for further relevant studies. The reference lists of the industry submissions were also scrutinised for additional studies. Due to resource limitations the search was restricted to English language papers only. All references were managed using Reference Manager (Professional Edition Version 11; Thomson ISI ResearchSoft) and Microsoft Access 2003 software.

Relevant studies were identified in two stages. Titles and abstracts returned by the search strategy were examined independently by two researchers (GR and MB) and screened for possible inclusion. Disagreements were resolved by discussion. Full texts of the identified studies were obtained. Two researchers (GR and MB) examined these independently for inclusion or exclusion, and disagreements were again resolved by discussion.

4.1.2. Inclusion and exclusion criteria

4.1.2.1. Study design

Inclusion criteria

For the review of clinical effectiveness, only systematic reviews of RCTs and RCTs were considered. The review protocol made provision for broadening search criteria to include

some observational evidence if insufficient systematic reviews or RCTs were identified; however, this proved unnecessary in view of the reasonable yield of evidence of a preferred design (see below).

Systematic reviews were used as a source for finding further RCTs and to compare with our systematic review. For the purpose of this review, a systematic review[71-73] was defined as one that has:

- A focused research question
- Explicit search criteria that are available to review, either in the document or on application
- Explicit inclusion/exclusion criteria, defining the population(s), intervention(s), comparator(s), and outcome(s) of interest
- A critical appraisal of included studies, including consideration of internal and external validity of the research
- A synthesis of the included evidence, whether narrative or quantitative.

Exclusion criteria

Studies were excluded if they did not match the inclusion criteria, and in particular:

- Non-randomised studies (except for AEs)
- Animal models
- Pre-clinical and biological studies
- Narrative reviews, editorials, opinions
- Non-English language papers
- Reports published as meeting abstracts only, where insufficient methodological details were reported to allow critical appraisal of study quality.

4.1.2.2. Population

Studies were included if they reported a population comprising adults with AD. Following the 2004 review, where trials included participants with mixed dementia, these were included where the predominant dementia was AD.

Participants in included trials were required to meet the definitions of disease severity specified in the technologies' UK marketing authorisations (MMSE 26–10 for donepezil, galantamine, and rivastigmine; MMSE 20–0 for memantine).

The exact inclusion criterion adopted for MMSE scores was defined as an approximation of the principle that at least 80% of a study's participants should be within the specified range. This approach relied on the assumption that reported baseline MMSE scores were normally distributed. On this basis, studies were included if the predefined thresholds were not exceeded by the reported mean baseline MMSE score ±0.8416SD, where 0.8416 is the inverse of the standard normal distribution corresponding to a probability of 0.8.

4.1.2.3. Interventions and comparators

Studies were included if the technologies they assessed fulfilled the following criteria:

Interventions: The four technologies under review were considered within their UK marketing authorisations:

- mild-to-moderately severe AD (measured by the MMSE 26–10): donepezil, galantamine, and rivastigmine
- moderate-to-severe AD (measured by the MMSE 20–0): memantine

Comparators: For people with **mild** AD the comparators of interest were placebo and/or best supportive care (i.e. treatment without AChEIs and without memantine). For people with **moderate** AD the comparators were donepezil, galantamine, rivastigmine, memantine, and placebo and/or BSC (i.e. treatment without AChEIs). For people with **severe** AD the comparator was treatment without memantine.

4.1.2.4. Outcomes

Studies were included if they reported data on one or more of the following outcomes:

- Measures of severity and response to treatment
- Behavioural symptoms
- Mortality
- Ability to remain independent
- Likelihood of admission to residential/nursing care

- HRQL of patients and carers
- AEs of treatment

4.1.3. Data extraction strategy

Data were extracted by GR into forms in bespoke software and checked by MB. Disagreements were resolved by discussion. The items extracted can be found in the data extraction forms of included studies which are available in Appendix 3.

4.1.4. Critical appraisal strategy

Assessments of study quality were performed according to the instrument developed for the 2004 review (which was based on criteria recommended by the NHS CRD.[71]). The instrument is summarised below; for full details, see Appendix 5 of the 2004 review.[2] Results were tabulated and the relevant aspects described in the data extraction forms.

4.1.4.1. Internal validity

The instrument sought to assess the following considerations:

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 point estimates and a measure of variability presented for the primary outcome measure?
9. Did the analyses include an intention-to-treat (ITT) analysis?
10. Were withdrawals and dropouts completely described?

In addition, methodological notes were made for each included study, including the reviewer's observation on: sample size and power calculations; participant attrition; methods of data analysis; and conflicts of interest.

4.1.4.2. External validity

External validity was judged according to the ability of a reader to consider the applicability of findings to a patient group and service setting. Study findings can only be generalisable if they describe a cohort that is representative of the affected population at large. Studies that appeared representative of the UK AD population with regard to these considerations were judged to be externally valid.

4.1.5. Methods of quantitative synthesis

Where data permitted, the results of individual trials were pooled using the methods described in the following section.

1.1.1.1. Pairwise meta-analysis

We used random-effects meta-analyses (DerSimonian and Laird model[74]) only, regardless of any statistical evidence of inter-study homogeneity. Heterogeneity was explored by visualisation of results and, in statistical terms, both Cochran's *Q* (compared to a χ^2 distribution)[75] and the I^2 statistic.[76;77] Small-study effects (including publication bias) were visualised using funnel plots and quantified using Egger's test.[78] (Appendix 4) Analyses were conducted using bespoke software, written in Visual Basic for Applications and applied in both Microsoft Access and Microsoft Excel. STATA 10.1 was used to generate forest plots (`metan` command) and to assess small-study effects (`metabias` command).

Where more than one arm of a contributing trial was relevant to any analysis, data were pooled to form a single meta-arm as the unit of analysis, as recommended in the Cochrane Handbook for Systematic Reviews of Interventions (¶16.5.4).[79] For the continuous outcome measures reported in this review, the mean for the combined arm is estimated as the weighted mean from the multiple separate arms (where the numbers in each arm provide the weights), and the standard deviation for the combined arm is calculated according to the usual formula:

$$s_C = \sqrt{\frac{\sum_{i=1}^{k}(n_i - 1)s_i^2}{\sum_{i=1}^{k}(n_i - 1)}}\ , \qquad \textbf{(1)}$$

where *i* indexes a total of *k* arms being combined, n_i is the number of participants in each arm, and s_i is the standard deviation for that arm.

All meta-analyses were stratified according to measurement population. Where multiple measurement populations were reported in an individual study, we used the highest-ranking according to a pre-specified hierarchy:

(i) true ITT;

(ii) last observation carried forward (LOCF);

(iii) observed cases only (OC).

The issue of how to deal with missing data point from drop outs in ITT analysis of dementia patients is a contentious one. Due to the natural course of this degenerative disease, the assumption of LOCF that disease progression stops at the last data point clearly does not reflect reality. Similarly, to use OCs only (i.e. not estimating any data points after drop out), may give misleading results. A better solution may be to apply the rate of decline found in the control group to all drop outs.[80]

We performed separate analyses for different periods of follow-up. The two lengths of follow-up for which data were generally available were approximately three months (12–16 weeks of treatment) and approximately six months (21–28 weeks) (figures showing these results are in the body of the text).

Where different dosages of drugs were found in various studies, we meta-analysed comparable groups separately (figures for commonly used doses in the UK are in the body of the text). We also performed a single analysis in which all dosages were combined (figures from these analyses are in Appendix 5). For continuous outcomes measured over a longitudinal period of follow-up, it is possible for investigators to report outcomes in two ways: the mean of each participant's observed change from a measured baseline score ("mean

change from baseline") or absolute measurements at the relevant juncture ("absolute value"). If randomisation is adequate, the difference between these values should be the same (i.e. the mean of the differences will be the same as the difference in the means). However, the dispersion of each measure may vary. It is stated in the Cochrane Handbook for Systematic Reviews of Interventions (¶9.4.5.2) that "[t]here is no statistical reason why studies with change-from-baseline outcomes should not be combined in a meta-analysis with studies with final measurement outcomes".[79] However, exploratory analyses showed that the inclusion of both types of data led to large differences in the results of meta-analyses, although this may be because the studies that only report final measurement data tend to be of a lower methodological standard (and, therefore, may also be more susceptible to biases that would distort reported treatment effect). As a result, we were not prepared to pool the two types of measurement, and all our meta-analyses rely on studies reporting mean change from baseline only.

4.1.5.1. Pooling of multiple outcome measures

In addition to pairwise meta-analyses of treatment effect pooled on each outcome's natural scale (weighted mean difference), we combined outcomes in a series of broad domains – cognitive, functional, behavioural, and global – to investigate the overall characteristics of reported effectiveness evidence in each area (figures in the body of the text and data sets used in the meta-analysis of pooled multiple outcome measures in Appendix 6).

In order to combine studies using different outcome measures within each domain, effect sizes were expressed as a standardised mean difference (SMD). The SMD expresses the size of the treatment effect in each trial relative to the variability observed in that trial. Accordingly, for a given trial *i*,

$$d_i = \frac{m_{1i} - m_{2i}}{s_i} \quad , \qquad \textbf{(2)}$$

where m_{1i} and m_{2i} represent the reported means in active treatment and control cohorts, respectively, and s_i is the pooled standard deviation across both groups, estimated as

$$s_i = \sqrt{\frac{(n_{1i}-1)SD_{1i}^2 + (n_{2i}-1)SD_{2i}^2}{N_i - 2}} \quad , \qquad \textbf{(3)}$$

where n_{1i}, n_{2i} and N_i represent the sample sizes of treated, control and combined cohorts, respectively, and the reported standard deviations of measurements in treated and control groups are SD_{1i} and SD_{2i}. In order to pool SMDs, it is necessary to derive the standard error, which is estimated as follows:

$$SE(d_i) = \sqrt{\frac{N_i}{n_{1i}n_{2i}} + \frac{d_i^2}{2(N_i - 2)}} \quad , \qquad \textbf{(4)}$$

The method assumes that the differences in standard deviations (SDs) among studies reflect differences in measurement scales and not real differences in variability among study populations.

Where studies reported more than one outcome contributing to the same domain, a weighted average of all SMDs was calculated, using the precision of the estimates as the weighting factor (this could be seen as a sub–meta-analysis, adopting a fixed-effects model with inverse variance weighting). So that such studies were not given spurious weight, the sample size for each outcome measure was divided by the total number of outcomes.

This approach has the advantage of enabling a broader evidence-base to be combined, but it has the disadvantage of requiring estimates to be pooled on a scale which has no direct clinical meaning. Accordingly, we used these analyses solely to explore the characteristics of the evidence-base, and not to draw direct conclusions about the magnitude of relative effectiveness of the comparators. In particular, we used the analyses as a basis for meta-regression (see below), and for assessing small-study effects.

4.1.5.2. Meta-regression

Where there was sufficient evidence (at least five individual datapoints in a meta-analysis), study level regression ("meta-regression") was used to explore the statistical heterogeneity across studies. Three prespecified covariates were explored: population age, population

sex, and baseline disease severity (as measured by MMSE). Because of inconsistencies in the evidence-base, it was not possible to undertake multivariate analyses, so regressions were conducted solely on a univariate basis. Meta-regression was undertaken in STATA 10.1 (`metareg` command), using the restricted maximum likelihood estimator, as recommended.[81;82] These figures are in Appendix 7 .

4.1.5.3. Mixed treatment comparison – indirect comparison

In addition to pairwise meta-analyses, where sufficient data was available, we synthesised information on all technologies and their comparators simultaneously, in a mixed treatment comparison (MTC) using Bayesian Markov Chain Monte-Carlo (MCMC) sampling.[83-86] The analyses were performed using WinBUGS 1.4.1; model code is reproduced in Appendix 8.

Vague prior distributions were used in the analyses (Normal[0, 0.000001] for mean difference between treatments; Uniform[0,2] for SD of random effects distribution). Point estimates and 95% credible intervals were calculated from 100,000 simulated draws from the posterior distribution after a burn-in of 10,000 iterations.

Outputs are presented in terms of treatment effect compared to a common baseline. In each case in the presented analyses, the available evidence networks included at least one placebo arm; therefore, the baseline treatment is always placebo. This is helpful, as it enables all MTC outputs to be interpreted on a common level. In addition to treatment effect relative to placebo, the posterior probability that each treatment is most effective is presented, simply calculated as the proportion of MCMC trials in which the given treatment had the highest (or lowest, for negative scales) estimated treatment of all comparators.

This approach assumes "exchangeability" of treatment effect across all included trials, such that the observed treatment effect for any comparison could have been expected to arise if it had been measured in the populations reported in all other included trials. Exchangeability was judged through examination of the trial populations and comparability of outcomes in the common treatment group facilitating the comparison. Figures representing these analyses are in the body of the text.

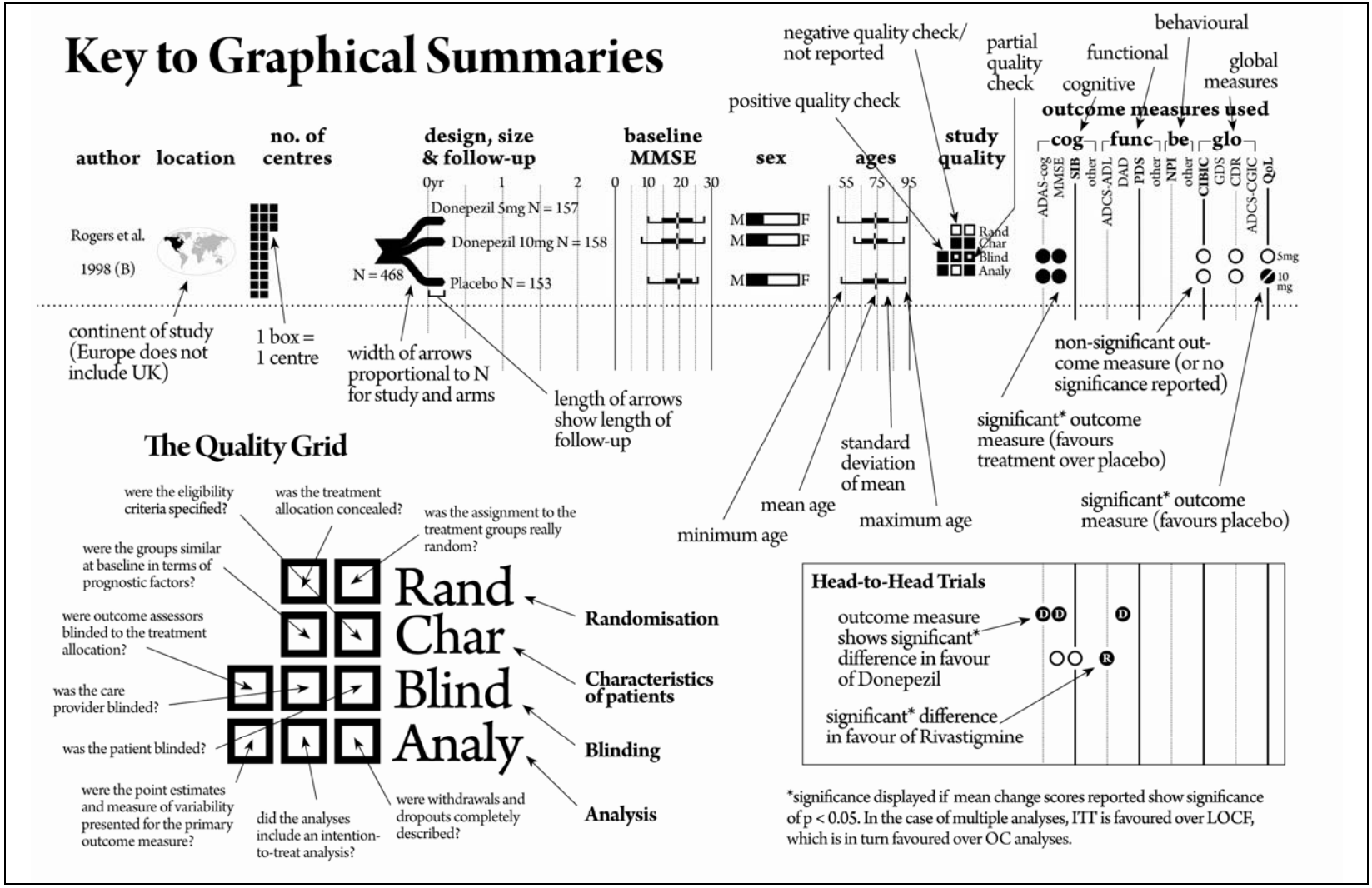
As for pairwise syntheses, we generated separate MTCs for different periods of follow-up (12–16 weeks and 21–28 weeks). We also generated separate analyses according to measurement population: LOCF only; ITT+LOCF; OC only; and all measurement populations

combined (Appendix 9). Where multiple measurement populations were reported in an individual study and more than one was pertinent to one of these analyses, we used the highest-ranking according to the hierarchy given above. Multiple relevant arms within a single study were pooled according to the methods detailed in ¶1.1.1.1, above, before being entered into the MTC.

4.1.6. Graphical representation of summary trial information

We present a novel approach to summarizing the complex information relating to each trial at the end of each comparison section. These figures graphically represent the location, size, MMSE score at baseline, gender, age, study quality and results in a format that allows quick comparison between trials. A key to understanding the graphics is presented below in *Figure 5*.

FIGURE 5 *Key to graphical summaries*



4.2. Results of the systematic review: Identification of evidence

From screening the titles and abstracts of the 1,843 references identified by our searches and additional sources, we retrieved 191 papers for detailed consideration, of which 21 were judged to meet the inclusion criteria for the review. The process is illustrated in detail in *Figure 6*. In assessing titles and abstracts, agreement between the two reviewers was moderately good (κ=0.642). At the full-text stage, agreement was moderate (κ=0.538). At both stages, initial disagreements were easily resolved by consensus.

The submissions from Eisai/Pfizer and Lundbeck contained a number of published and unpublished items that we have excluded from our review because they did not meet our inclusion criteria. A list of these items with reasons for their exclusion can be found in Appendix 10. A list of ongoing trials can be found in Appendix 11.

4.3. Results: systematic reviews

Our searches found four systematic reviews which met our inclusion criteria. These were critically appraised using the Preferred Reporting Items for Systematic Reviews and Meta-analyses (PRISMA) statement checklist, which describes 27 items that a report of a systematic review or meta-analysis should contain.[72] A summary table of whether these quality indicators were present in these systematic reviews can be found in Appendix 12. The references of each systematic review were checked to see if they held any additional includable trials, no further includable studies were found. A brief summary of each systematic review can be seen below.

4.3.1. Summary of included systematic reviews and meta-analyses

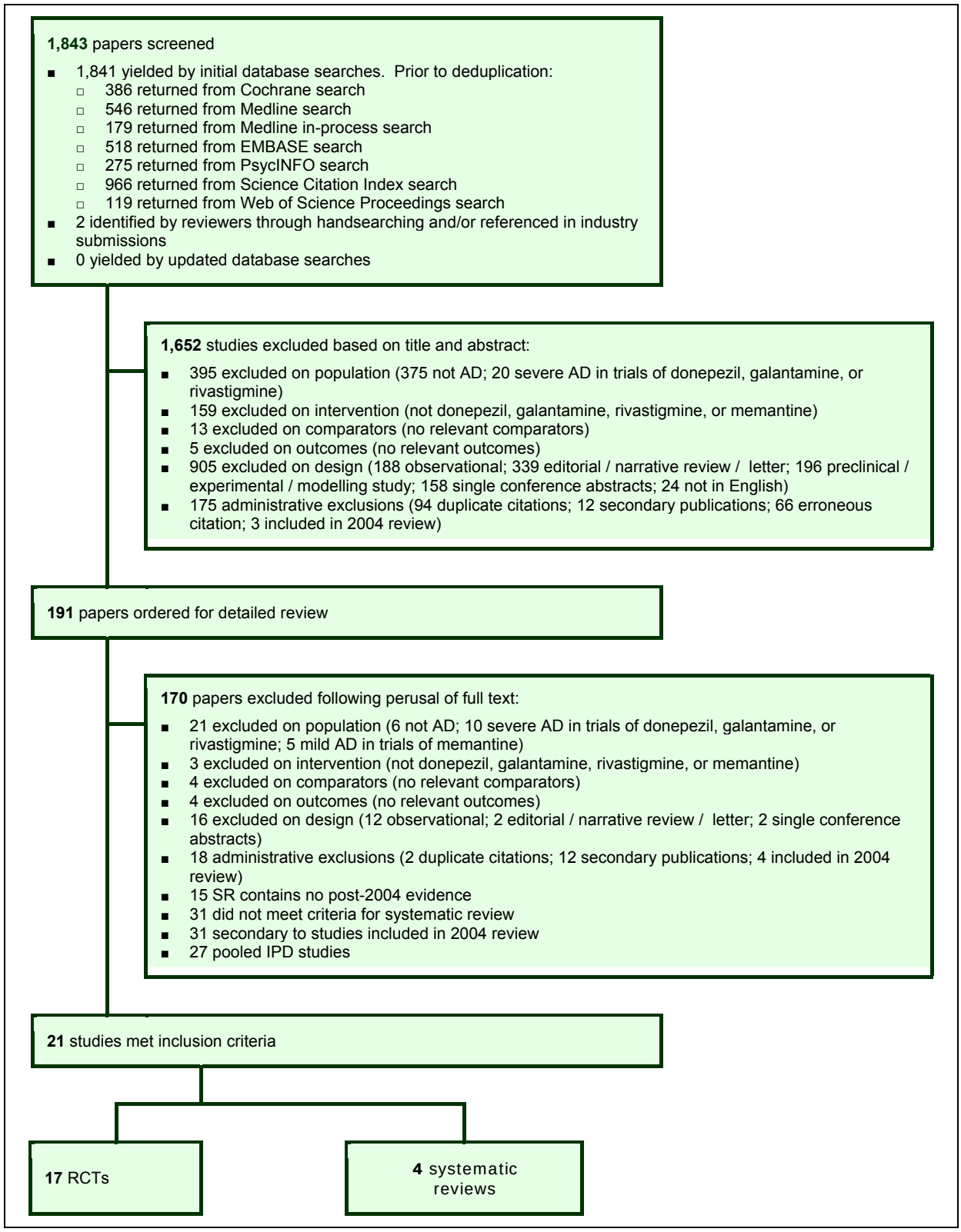
Donepezil

No systematic reviews of donepezil were found that matched our inclusion criteria.

Galantamine

No systematic reviews of donepezil were found that matched our inclusion criteria.

FIGURE 6 *Identification of published evidence for review*



Rivastigmine

Birks and colleagues (2009)[87] conducted a Cochrane review of rivastigmine compared with placebo for people with mild-to-moderate AD. They found nine trials with a total of 4,775 participants. The review found that the use of high doses of rivastigmine (6 to 12mg daily) was associated with a two point improvement on the ADAS-Cog compared to placebo (ITT WMD -1.99 [95% CI -2.49 to -1.50], and a 2.2 point improvement on the Progressive Deterioration Scale (PDS) for ADL, (ITT weighted mean difference -2.15 [95% CI -3.16 to-1.13] at 26 weeks. The authors concluded that rivastigmine gave benefit to people with mild-to-moderate AD when compared to placebo. The review also considered delivery of the drug by transdermal patch. It found that the lower dose patch (9.6 mg/day) was associated with fewer side-effects than the capsules or the higher dose patch (17.4 mg/day), and produced similar efficacy. The main AEs were gastrointestinal (nausea and vomiting) and usually occurred during the titration phase.

All cholinesterase inhibitors

The German Institute for Quality and Efficiency in Health Care (IQWiG) conducted a systematic review and meta-analysis of all the cholinesterase inhibitors included in this report for people with mild-to-moderate AD.[88] They included RCTs up to June 2006 in their systematic review, and found 27 studies with a total of 9,883 participants. Only four of these trials met our inclusion criteria.[89-92] The IQWiG concluded that all the AChEis provided benefit in improving or maintaining cognitive function and ADLs, and that galantamine alleviated psychological symptoms. However, none of the studies provided evidence of improvement in QoL. A summary table of these results can be found in Appendix 13.

Memantine

A systematic review of memantine for dementia was carried out by Raina and colleagues (2008)[93] Their inclusion criteria were broader than ours and included all major types of dementia, patients with mild to moderate disease severity and all drugs for treating dementia. Of the 59 studies they included only two met our inclusion criteria for trials.[89;94] The data syntheses from this systematic review are not relevant to this technology assessment report and will not be discussed.

All included drugs

Hansen and colleagues (2007)[95] conducted a systematic review and meta-analysis of functional outcomes from the use of donepezil, galantamine, rivastigmine and memantine for people with mild to moderate AD. They included 13 RCTs, 12 of which were included in the previous TAR (111). The new study, which is included in this review, is Brodaty and colleagues (2005)[89] Overall they found a small effect size (*d*=0.1-0.4) favouring drug treatment. A meta-regression showed that this effect was not affected by disease severity, age, gender and drug dose. AEs were most commonly gastrointestinal.

4.4. Manufacturers' submissions on clinical effectiveness

Three reviews were presented summarising evidence on the effectiveness of donepezil, galantamine and memantine by the manufacturers of each of the drugs. Although not part of the PenTAG systematic review they are presented here for convenience and because their findings are compared with our own review. Each submission is briefly discussed in the sections below.

4.4.1. Donepezil

Eisai Ltd and Pfizer Ltd submitted a systematic review as part of their joint submission on donepezil. It included both RCTs and targeted non-RCT/observational studies. Concerning the effect of donepezil relative to placebo the reported results of effect on cognition, function, behaviour and global impact were consistent with the results of the PenTAG review. There was however limited information on any summary estimates of effect in the manufacturer submission. Challenges to the validity of the AD2000 trial were re-emphasised.

Published meta-analyses were used to explore whether the effect of donepezil varied depending on the severity of AD, particularly the effectiveness in patients with mild AD. These suggested that a beneficial effect of donepezil relative to placebo on cognition, global impact and behaviour was present for patients with mild AD. The summary estimates quoted for mild AD were broadly similar to the overall summary estimates calculated in the PenTAG systematic review.

Results from non-RCT and observational data were presented to support the following additional aspects of the effectiveness of donepezil:

- duration of effectiveness extending beyond 6 months up to at least 3 years
- worsening of symptoms following withdrawal of treatment
- emergence of benefit after initial absence of changes suggesting response
- impact on carers particularly care-giver stress and carer time
- trends towards reductions in anti-psychotic medication use
- reductions in mortality

4.4.2. Galantamine

Shire Pharmaceuticals presented a summary of all available RCTs (not just those from 2004 onwards) comparing galantamine with placebo, but did not indicate how the review had been conducted. They emphasised the importance of newer dosing regimens and highlighted deficiencies in the previous systematic review by SHTAC, particularly concerning failure to include a study directly comparing galantamine with donepezil

The pooled summary estimates presented for the effect of galantamine on cognition, behaviour and function were consistent with the summary estimates in the PenTAG systematic review. The Shire submission provided additional analyses indicating an increase in effect with increasing severity of disease. Similar analyses could not be done in the PenTAG systematic review because of the requirement for individual patient data.

4.4.3. Memantine

Lundbeck presented a meta-analysis of pivotal trials as part of their submission. Although some details on the methods of analysis were provided, there was no information on how the pivotal trials were ascertained. The inclusion criteria were given and in essence the included studies were double-blind RCTs comparing memantine with placebo measuring cognition, disability, global health state and behaviour at three or six months. The need for individual patient data was further stipulated to allow sub-group analysis. There were six included studies in the main analysis covering all periods, not just the 2004 onwards. The reasons why some studies were included in the Lundbeck analysis but not included in the PenTAG meta-analysis are documented in Appendix 10. Briefly, these were that Lundbeck's pooling methods relied on the availability of IPD to which PenTAG did not have access and Lundbeck were prepared to pool data from trials of memantine + ACHEIs v. ACHEIs alone with data from trials of memantine monotherapy v. placebo to produce a single estimate of

memantine effect. PenTAG were not comfortable with the assumptions necessary to justify such a single analysis. Not withstanding this, the direction and size of effect of memantine relative to placebo on cognition, disability, global health state and behaviour are consistent between the Lundbeck and PenTAG analysis. In this, account needs to be taken that the results in the Lundbeck submission are presented as SMD whereas those in the PenTAG analysis were WMDs. Approximate interconversion is achieved by multiplying or dividing by the pooled SD. The 95% CI are narrower in the Lundbeck analysis because of the greater number of included studies. The submission identified no evidence that the effectiveness varied by severity of Alzheimer's disease, by past use of AChEIs or by concurrent use AChEIs. These analyses could not be repeated in the PenTAG systematic review because they depend on individual patient data.

Lundbeck also examined whether there was evidence of a difference in effectiveness depending on the presence of Agitation/Aggression and/or Psychotic Symptoms (APS), defined by the baseline NPI score being ≥3 (as opposed to the definition of >0 used in the last the submission for the last NICE guidance). The results suggested that there is greater effectiveness in patients with APS but again these analyses could not be repeated in the PenTAG systematic review because they depend on individual patient data.

4.5. Unavailable evidence

Subgroup analyses

The study protocol specified that if evidence allowed subgroups based on disease severity, response to treatment, behavioural disturbance and comorbidities should be considered. However, none of the included trials reported any of these subgroup analyses. Therefore, we are unable to comment on them.

Outcomes

None of the included trials reported mortality or institutionalisation outcomes, or reported on outcomes beyond 28 weeks.

4.6. Results: pairwise comparisons

4.6.1. Donepezil *v.* placebo

4.6.1.1. Identified evidence

The 2004 review identified 14 RCTs investigating the effectiveness of donepezil compared with placebo, those reported by AD2000 (2004),[96] Burns and colleagues. (1999),[97] Gauthier and colleagues. (2002),[98] Greenberg and colleagues. (2000),[99] Holmes and colleagues. (2004),[100] Homma and colleagues. (2000),[101] Krishnan and colleagues. (2003),[102] Mohs and colleagues. (2001),[103] Nunez and colleagues. (2003)[104;105] (NB this trial was reviewed in poster form in 2004; a full publication, authored by Johannsen and colleagues. (2006) is now available[105], from which we have extracted data; however, for consistency with the 2004 review, we continue to refer to this RCT as Nunez and colleagues. (2003), Rogers and colleagues. (1998),[106] Rogers and colleagues. (1998),[107] Rogers & Friedhoff. (1996),[108] Seltzer and colleagues. (2004) (NB this trial was reviewed on a CiC basis using information supplied by the manufacturer in 2004; a full publication, authored by Seltzer and colleagues. (2004) is now available[109]), and Winblad and colleagues. (2001)[110] (with additional information contained in Wimo and colleagues. (2003)[111]).

Our searches identified an additional five RCTs. These are, Mazza and colleagues (2006),[112] Moraes and collegues (2006),[113] Moraes and colleagues (2008),[114] Peng and colleagues (2005),[115] and Winstein and colleagues (2007).[116] A summary of their design characteristics can be found in *Table 3* and the interventions, comparators and baseline characteristics of the participants in *Table 4*. Critical appraisal of these small studies showed that none were of good quality; neither was reporting adequate randomisation nor allocation concealment. A summary of the markers of internal validity is presented in *Table 5*.

TABLE 3 *Design of included studies – donepezil* v. *placebo*

Study details	Inclusion criteria	Exclusion criteria	Methodological notes	Other
Mazza et al. (2006)[112] **Design:** Parallel double-blind RCT **Country:** Italy? **No. of centres:** 1 **No. randomised:** 76 **Maximum follow-up:** 24 **MMSE range included:** 13–25	AD (DSM-IV criteria) Brief Cognitive Rating scale mean score 3-5 Hachinski Iscaemic Score <4 Adequate level of premorbid intelligence (IG>80, global assessment)	Dementia of other aetiology Severe organic diseases (tumours, severe infectious diseases, brain trauma, epilepsy, cerebrovascular malformations, alcohol or drug abuse) Pseudodementia or a histiory of schizophrenic or affective psychoses (Geriatric Depression Scale, 15-item version, total score <9) Vasoactive drugs, nootropics and long-term treatment with other drugs were proscribed during the study, with the exception of low doses of benzodiazepines and neuroleptics in the treatment of behavioural disturbances.	**Sample attrition / dropout:** 60 of 76 randomised patients completed the study (a further 41 were excluded during the run-in period). Reasons for drop-out not reported. **Randomisation and allocation:** Randomisation computer-generated (whether unreadable before allocation is not stated). Appearance of pills and placebo not reported. **Power calculation:** Not reported	**Therapy common to all participants:** Single-blind placebo 4-week run-in period (in order to exclude placebo responders) **Study Funding:** Not reported **Other conflicts:** Not reported
Moraes et al. (2006)[113] **Design:** Parallel double-blind RCT **Country:** Brazil **No. of centres:** 1 **No. randomised:** 35 **Maximum follow-up:** 26 **MMSE range included:** not reported	Probable AD (AD and Related Disorders Association criteria) Clinical Dementia Rating (Brazilian version) 1-2 (mild to moderate)	Other causes of dementia Other current severe medical or psychiatric disease Evidence of moderate to severe sleep disorders, based on medical, sleep, and psychiatric interviews Apnoea-hypoapnoea index >10/h and periodic leg movement index >5/h at baseline polysomnographic recording Psychoactive drugs in the month prior to entering the study	**Sample attrition / dropout:** 8 patients left the study due to technical difficulties in polysomnography recordings **Randomisation and allocation:** Randomisation process not reported. Individual responsible for the random allocation of patinets to the trial arms was blind to the treatment code (how blinding was attained is not reported). Appearance of donepezil and placebo tablets is not described. **Power calculation:** Data from 10 patients was initially analysed for sample size estimation (procedure not reported). Based on this analysis, a sample size of 15 subjects in each group was calculated to set out a difference of 8 percentage points in REM sleep percentage (significance level of 1% and power of 95%). To assess the interaction term in the ANOVA model, 27 subjects were required in each group (sample size not attained) – power of 80% was possible with the sample size analysed.	**Therapy common to all participants:** 2 nights of polysomnographic recording (for purposes of habituation) **Study Funding:** FAPESP (Fundacao de Amparoa Pesquisa do Estado de Sao Paulo) AFIP (Associacao Fundo de Incentivo a Psicofarmacolgia) **Other conflicts:** Authors state no financial conflicts of interest. No financial support from industry for study.

Study details	Inclusion criteria	Exclusion criteria	Methodological notes	Other
Moraes et al. (2008)[114] **Design:** Parallel double-blind RCT **Country:** Brazil **No. of centres:** 1 **No. randomised:** 23 **Maximum follow-up:** 12 **MMSE range included:** 6–27	AD (ADRDA criteria) Rating of 1-2 (mild to moderate) on Brazilian version of Clinical Dementia Rating	Rating of >=3 on Brazilian version of Clinical Dementia Rating Other causes of dementia Other current severe medical or psychiatric disease Psychoactive drugs in the month prior to entering the study	**Sample attrition / dropout:** Not reported **Randomisation and allocation:** Randomisation performed using computer-generated random number list (0-1) with uniform distribution, with patients consecutively allocated to the two treatment groups (<=0.5 to group A, >0.5 to group B). Donepezil and placebo pills were 'packed in the same fashion', but precise appearance of pills not reported. **Power calculation:** Not reported	**Therapy common to all participants:** 2 nights of polysomnographic recording (for purposes of habituation) **Study Funding:** FAPESP (Fundacao de Amparoa Pesquisa do Estado de Sao Paulo) AFIP (Associacao Fundo de Incentivo a Psicofarmacolgia) **Other conflicts:** Authors state no conflicts of interest to disclose
Peng et al. (2005)[115] **Design:** Parallel double-blind RCT **Country:** China **No. of centres:** 15 hospitals in Beijing, Shanghai, and Guangzhou **No. randomised:** 90 **Maximum follow-up:** 12 **MMSE range included:** 10–24	AD (NINCDS-ADRDA and DSM-IVR criteria) >=55y old In female patients, menopause >=2y Sufficinet vision and hearing to complete assessments	Other disease that may lead to dementia Severe heart or kidney dysfunction, active peptic ulcer, or active epilepsy Allergy to cholinergic drugs	**Sample attrition / dropout:** 89 of 90 completed the study.n=1 dropped out due to adverse event (dizziness) **Randomisation and allocation:** Randomisation procedure not described. Placebo described as having the same colour, shape, flavour and size as donezepil **Power calculation:** Not reported	**Therapy common to all participants:** None **Study Funding:** Not reported **Other conflicts:** Not reported

Study details	Inclusion criteria	Exclusion criteria	Methodological notes	Other
Winstein et al. (2007)[116] **Design:** Parallel double-blind RCT **Country:** USA **No. of centres:** 1 **No. randomised:** 10 **Maximum follow-up:** 4 **MMSE range included:** 11–26	Probable AD diagnosis (criteria not reported) Independent in ambulation Alert Able to follow simple instructions	Delirium Familial tremor Parkinson's Disease Stroke Peripheral neuropathy Dementia due to other than probable AD Use of any concurrent pharmaceutical treatment for cognitive dysfunction	**Sample attrition / dropout:** 10 of 10 completed study **Randomisation and allocation:** Randomisation procedure not described. Placebo described as identical in appearance to donepezil. **Power calculation:** Not reported	**Therapy common to all participants:** None **Study Funding:** USC AD Research Centre, AD Research Centres of California, and Pfizer, Inc. **Other conflicts:** None reported

TABLE 4 *Interventions, comparators, and baseline characteristics of participants in included studies – donepezil* v. *placebo*

Study	Arm	Dose (mg/d)	Dosage details	N	Age	Sex (n male)	Race (n white)	Weight (kg)	Education (yrs)	Duration of dementia (mo)	ADAS-cog	MMSE
Moraes et al. (2006)[113]	Donepezil	5–10	Starting daily dose of 5mg for the first month, increased to 10mg/d in the second month	17	77.4 (SD 6.60)	4 (23.5%)			4.40 (SD 3.60)		35.6 (SD 13.7)	
	Placebo	-	Single daily dose	18	74.5 (SD 9.80)	7 (38.9%)			6.00 (SD 5.20)		39.0 (SD 18.5)	
Mazza et al. (2006)[112]	Donepezil	5	5mg daily	25	64.5 (SD 6.00)	13 (52.0%)						18.6 (SD 3.47)
	Placebo	-	Not reported	26	69.8 (SD 3.00)	10 (38.5%)						18.8 (SD 3.63)
Moraes et al. (2008)[114]	Donepezil	5–10	Single dose of 5mg (administered at bedtime) in the first month, increased to single dose of 10mg in second month	11	76.8 (SD 6.20)	3 (27.3%)					34.5 (SD 15.8)	19.0 (SD 3.60)
	Placebo	-	Single dose administered at bedtime	12	72.6 (SD 11.0)	5 (41.7%)					29.3 (SD 17.3)	17.2 (SD 7.80)
Peng et al. (2005)[115]	Donepezil	5	Same dose administered throughout duration of study	46	72.6 (SD 6.80)	21 (45.7%)						17.8 (SD 2.30)
	Placebo	-	-	43	71.8 (SD 8.20)	19 (44.2%)						18.2 (SD 2.70)
Winstein et al. (2007)[116]	Donepezil	5	One tablet taken nightly	5	84.2 (SD 8.67)	2 (40.0%)					24.0 (SD 3.08)	19.2 (SD 3.35)
	Placebo			5	88.0 (SD 7.62)	1 (20.0%)					26.0 (SD 11.6)	20.2 (SD 4.09)

TABLE 5 *Markers of internal validity of included studies – donepezil* v. *placebo*

	Was the assignment to the treatment groups really random?	Was the treatment allocation concealed?	Were the groups similar at baseline in terms of prognostic factors?	Were the eligibility criteria specified?	Were outcome assessors blinded to the treatment allocation?	Was the care provider blinded?	Was the patient blinded?	Were the point estimates and measure of variability presented for the primary outcome measure?	Did the analyses include an intention-to-treat analysis?	Were withdrawals and dropouts completely described?
Mazza *et al.* (2006)[112]	PARTIAL	INADEQUATE	REPORTED – YES	INADEQUATE	PARTIAL	PARTIAL	PARTIAL	ADEQUATE	PARTIAL	PARTIAL
Moraes *et al.* (2006)[113]	UNKNOWN	INADEQUATE	REPORTED – YES	INADEQUATE	PARTIAL	PARTIAL	PARTIAL	ADEQUATE	UNKNOWN	PARTIAL
Moraes *et al.* (2008)[114]	INADEQUATE	INADEQUATE	REPORTED – YES	UNKNOWN	PARTIAL	ADEQUATE	ADEQUATE	ADEQUATE	UNKNOWN	INADEQUATE
Peng *et al.* (2005)[115]	UNKNOWN	UNKNOWN	REPORTED – YES	UNKNOWN	UNKNOWN	ADEQUATE	ADEQUATE	ADEQUATE	INADEQUATE	ADEQUATE
Winstein *et al.* (2007)[116]	UNKNOWN	UNKNOWN	REPORTED – YES	INADEQUATE	UNKNOWN	ADEQUATE	ADEQUATE	INADEQUATE	PARTIAL	ADEQUATE

4.6.1.2. Evidence of clinical effectiveness

4.6.1.2.1. Cognition

In 2004, Loveman and colleagues summarised the evidence they found for donepezil v. placebo for cognitive outcomes as follows:

> *"Six RCTs showed that donepezil appears to confer a statistically significant benefit to participants on the ADAS-Cog scale when compared to placebo. The benefit varies according to the dose of donepezil with higher doses of donepezil tending to show increasing benefit. Because the mean change scores varied quite considerably between the included studies, this dose related trend can particularly be seen within individual trials, although no direct statistical comparisons were made in any of these. The mean change scores were however varied between the included studies. Eight RCTs showed trends towards better MMSE score in the donepezil treated groups when compared to the placebo groups, although this was not always demonstrated to be statistically significant. These trends were mirrored in one unpublished trial of people with mild AD."*[2]

New data

In the studies we found published since 2004, four showed significant cognitive benefit for donepezil v. placebo. However, only two of these trials[112;116] estimated the missing values from drop outs using ITT analysis. The others made estimates using OCs only, thus potentially magnifying any benefit from donepezil and biasing their results in favour of the intervention (Section 1.1.1.1.). A summary of the results from cognitive measures can be seen in *Table 6*.

TABLE 6 *Measures of cognition in included studies – donepezil* v. *placebo*

Study	Subgroup	Outcome	Type[a]	Donepezil		Placebo		P
				N	Mean	N	Mean	
Moraes et al. (2008)[114]	OC population	ADAS-cog – 13wk	A	11	29.7 (SD 15.7)	12	31.8 (SD 18.5)	<0.05[b]
Winstein et al. (2007)[116]	ITT population	ADAS-cog – 4wk	MC	5	-5 (SD 2)	5	0 (SD 4.85)	0.066[c]
		Serial Reaction Time Task – 4wk	MC	5	3.325 (SD 8.39)	5	1.65 (SD 10.1)	0.782[c]
Moraes et al. (2006)[113]	OC population	ADAS-cog – 13wk	A	17	30.7 (SD 13.9)	18	40.9 (SD 19.4)	0.085[c]
		ADAS-cog – 26wk	A	17	28.3 (SD 12.3)	18	42.8 (SD 18.7)	<0.01[d]
Mazza et al. (2006)[112]	ITT population	MMSE – 24wk	A	25	19.8 (SD 3.16)	26	18.6 (SD 3.66)	NS[e]
			MC	25	1.2 (SD 12.2)	26	-0.25 (SD 5)[f]	0.06[e]
		Syndrom Kurztest – 24wk	A	25	11.8 (SD 2.9)	26	16.9 (SD 3.9)	0.01[e]
			MC	25	-3.3 (SD -2.55)	26	0.9 (SD 1.3)	<0.001[e]
		CGI: item 2 (cognitive) – 24wk	A	25	3.6 (SD 0.94)	26	5.2 (SD 0.95)	0.01[e]
			MC	25	-0.9 (SD 1.02)	26	0.15 (SD 0.338)	<0.001[e]
Peng et al. (2005)[115]	OC population	MMSE – 12wk	A	46	22.1 (SD 2)	43	18.7 (SD 2.4)	<0.01[g]

[a] A=absolute value at specified juncture; MC=mean change from baseline at specified juncture

[b] ANOVA

[c] student's t-test (calculated by reviewer)

[d] two-way ANOVA, with treatment group and treatment time as the main factors

[e] ANOVA, covarying age, gender, and severity of cognitive impairment at baseline

[f] reported 95%CI is asymmetric, suggesting calculation error

[g] t-test

Synthesis with existing evidence-base

The data from the new trials were synthesized with those from Loveman and colleagues report [2] by random-effects meta-analysis. This was conducted considering ADAS-cog and then MMSE as the outcomes, measuring differences between donepezil (all doses) and placebo at 12 and 24 weeks post randomisation. The results can be seen in *Figure 7* and *Figure 8*. We also meta-analysed the data by 5 mg/d and all doses combined, these results can be found in Appendix 5. We then went onto explore the effect of pooling the entire cognitive outcome measures at 24–26 weeks, the results of this can be seen in *Figure 11*.

ADAS-cog

We found no new studies reporting the ADAS-cog at 12 or 24 weeks. The meta-analyses presented below are of studies included in the previous assessment report. The overall pooled estimates shows a benefit from donepezil compared to placebo that increases over time; 12 weeks WMD=-1.97 (95%CI -3.38, -0.56), p=0.006 and 24 weeks WMD=-2.90 (95%CI -3.61, -2.18), p<0.001. (*Figure 7* and *Figure 8*).

FIGURE 7 *Random-effects meta-analysis: ADAS-cog at 12wk (mean change from baseline) – donepezil (10mg/d)* v. *placebo*

	Donepezil			Placebo				
	N	mean	SD	N	mean	SD	WMD (95%CI)	Wght
LOCF analysis								
Rogers et al. (1998)[107]	155	-2.70	5.35	150	0.40	5.27	-3.100 (-4.292, -1.908)	34.1
Nunez et al. (2003)[104;105]	94	0.65	6.11	98	0.70	6.14	-0.050 (-1.782, 1.682)	26.9
subtotal (*Q*=8.08 [*p* on 1 d.f.=0.004]; I^2=87.6%; τ^2=4.076)							**-1.642 (-4.628, 1.344)** *p*=0.281	**61.0**
OC population								
Burns et al. (1999)[97]	273	-1.90	4.96	274	0.40	4.97	-2.300 (-3.132, -1.468)	39.0
subtotal							**-2.300 (-3.132, -1.468)** *p*<0.001	**39.0**
Overall pooled estimate (*Q*=8.16 [*p* on 2 d.f.=0.017]; I^2=75.5%; τ^2=1.147) Inter-stratum heterogeneity: *p*=0.784 Small-study effects: Egger's *p*=0.586							**-1.969 (-3.379, -0.559)** *p*=0.006	

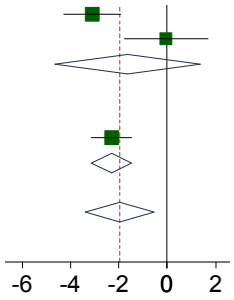
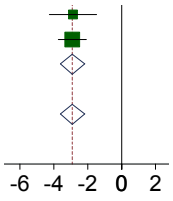


FIGURE 8 *Random-effects meta-analysis: ADAS-cog at 24wk (mean change from baseline) – donepezil (10mg/d)* v. *placebo*

	Donepezil			Placebo				
	N	mean	SD	N	mean	SD	WMD (95%CI)	Wght
LOCF analysis								
Rogers et al. (1998)[106]	150	-1.06	6.25	153	1.82	6.06	-2.880 (-4.266, -1.494)	26.5
Burns et al. (1999)[97]	273	-1.20	4.96	274	1.70	4.97	-2.900 (-3.732, -2.068)	73.5
subtotal (*Q*=0.0 [*p* on 1 d.f.=0.981]; I^2=0.0%; τ^2=0.000)							**-2.895 (-3.608, -2.182)** *p*<0.001	**100.0**
Overall pooled estimate (*Q*=0.0 [*p* on 1 d.f.=0.981]; I^2=0.0%; τ^2=0.000) Small-study effects: not calculable							**-2.895 (-3.608, -2.182)** *p*<0.001	



MMSE

Similarly, no new evidence was found for the outcome measure MMSE at 12 weeks post-randomisation but one new study was found with measures at 24 weeks follow-up. The meta-analyses below show an overall benefit from donepezil v. placebo when measured on the MMSE. 12 weeks (10 mg/d) WMD=-1.17 (95%CI 0.88, 1.45), p<0.001 and 24 weeks (5mg/d and10mg/d) WMD=1.21 (95%CI 0.84, 1.57), p<0.001. (*Figure 9* and *Figure 10*).

FIGURE 9 *Random-effects meta-analysis: MMSE at 12wk (mean change from baseline) – donepezil (10mg/d)* v. *placebo*

	Donepezil			Placebo				
	N	mean	SD	N	mean	SD	WMD (95%CI)	Wght
ITT population								
AD2000 (2004)[96]	245	0.93	3.24	263	0.00	2.96	0.930 (0.389, 1.471)	26.9
subtotal							**0.930 (0.389, 1.471)**	**26.9**
							p<0.001	
LOCF analysis								
Rogers et al. (1998)[107]	156	1.30	3.00	150	0.04	3.06	1.260 (0.581, 1.939)	17.1
Nunez et al. (2003)[104;105]	93	1.41	3.18	99	0.58	3.18	0.830 (-0.071, 1.731)	9.7
Holmes et al. (2004)[100]	41	-0.10	3.84	55	-1.80	3.71	1.700 (0.169, 3.231)	3.4
subtotal (Q=1.08 [p on 2 d.f.=0.584]; I^2=0.0%; τ^2=0.000)							**1.171 (0.659, 1.682)**	**30.1**
							p<0.001	
OC population								
Mohs et al. (2001)[103]	171	1.45	3.43	178	-0.15	3.34	1.600 (0.889, 2.311)	15.6
Winblad et al. (2001)[110]	127	0.69	2.59	128	-0.11	3.28	0.800 (0.075, 1.525)	15.0
Gauthier et al. (2002)[98]	84	2.00	4.12	96	0.00	3.92	2.000 (0.820, 3.180)	5.7
Seltzer et al. (2004)[109]	79	1.58	3.33	51	0.40	2.86	1.175 (0.100, 2.250)	6.8
subtotal (Q=3.91 [p on 3 d.f.=0.271]; I^2=23.3%; τ^2=0.062)							**1.322 (0.822, 1.823)**	**43.0**
							p<0.001	
Overall pooled estimate							**1.165 (0.884, 1.445)**	
(Q=6.14 [p on 7 d.f.=0.524]; I^2=0.0%; τ^2=0.000)							*p*<0.001	
Inter-stratum heterogeneity: p=0.563								
Small-study effects: Egger's p=0.213								

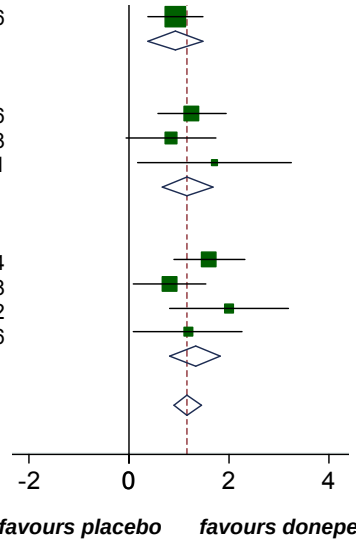
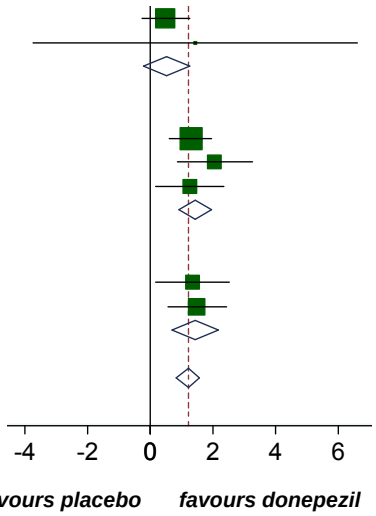


FIGURE 10 *Random-effects meta-analysis: MMSE at 24wk (mean change from baseline) – donepezil (all dosages)* v. *placebo*

	Donepezil			Placebo				
	N	mean	SD	N	mean	SD	WMD (95%CI)	Wght
ITT population								
AD2000 (2004)[96]	211	0.50	-	229	0.00	-	0.500[a] (-0.250, 1.250)	24.0
Mazza et al. (2006)[112]	25	1.20	12.25	26	-0.25	5.00	1.450 (-3.720, 6.620)	0.5
subtotal (Q=0.13 [p on 1 d.f.=0.722]; I^2=0.0%; τ^2=0.000)							**0.520 (-0.223, 1.262)**	**24.5**
							p=0.170	
LOCF analysis								
Rogers et al. (1998)[106]	303[b]	0.31	3.57	154	-0.97	3.47	1.284 (0.604, 1.964)	29.1
Gauthier et al. (2002)[98]	91	1.50	4.29	100	-0.56	4.00	2.060 (0.880, 3.240)	9.7
Seltzer et al. (2004)[109]	91	1.35	3.34	55	0.10	3.15	1.250 (0.171, 2.329)	11.6
subtotal (Q=1.38 [p on 2 d.f.=0.502]; I^2=0.0%; τ^2=0.000)							**1.425 (0.908, 1.943)**	**50.4**
							p<0.001	
OC population								
Mohs et al. (2001)[103]	111	1.80	4.21	96	0.45	4.29	1.350 (0.188, 2.512)	10.0
Winblad et al. (2001)[110]	121	0.40	3.74	120	-1.09	3.72	1.490 (0.548, 2.432)	15.2
subtotal (Q=0.03 [p on 1 d.f.=0.854]; I^2=0.0%; τ^2=0.000)							**1.434 (0.703, 2.166)**	**25.2**
							p<0.001	
Overall pooled estimate							**1.206 (0.839, 1.573)**	
(Q=5.89 [p on 6 d.f.=0.436]; I^2=0.0%; τ^2=0.000)							*p*<0.001	
Inter-stratum heterogeneity: p=0.114								
Small-study effects: Egger's p=0.459								



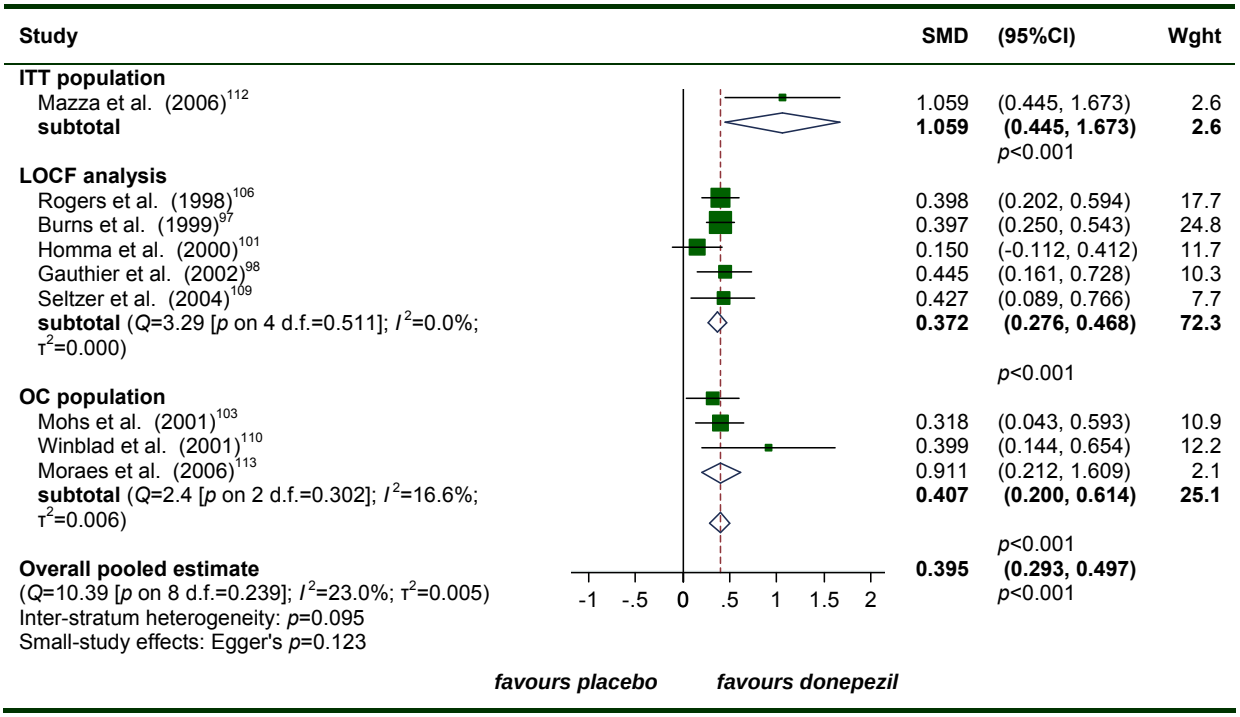
[a] WMD and error bars provided in publication; SE estimated on assumption that error-bars represent 95%CIs

[b] pooled 5mg/d and 10mg/d arms

4.6.1.2.1.1. Pooled multiple outcome measures

Two new studies were found to add to this combined meta-analysis of cognitive outcome measures at 24–26 weeks. The overall pooled estimate showed a significant cognitive benefit from donepezil compared to placebo, SMD=0.40 (95%CI 0.29, 0.50), p<0.001 (*Figure 11*). The data set used in this meta-analysis can be found in Appendix 6.

FIGURE 11 *Random effects meta analysis: cognitive outcomes (SMD) at 24–26wk – donepezil (all dosages) v. placebo*



4.6.1.2.2. Functional

The 2004 assessment report found that:

> *"A variety of functional measures were used in eight RCTs. Donepezil had some effect in improving or limiting further deterioration on ADLs when compared to placebo, but this was not always statistically significant, particularly over longer durations of follow-up. One trial reported time to loss of ADL and/or time to institutional care and found that donepezil conferred no advantage to placebo."*[2]

New data

We found only one new RCT measuring functional outcomes for this comparison. This small, poorly reported trial showed a significant benefit from donepezil (5 mg/d) for ADLs in an OC

measured population at 12-weeks follow-up, mean difference: I=40.5 (SD 7.6), C=49.5 (SD 6.3), p<0.01 (see *Table 7*).

TABLE 7 *Measures of functional ability in included studies – donepezil* v. *placebo*

Study	Subgroup	Outcome	Type	Donepezil		Placebo		*p*
				N	Mean	N	Mean	
Peng et al. (2005)[115]	OC population	ADL – 12wk	absolute value	46	40.5 (SD 7.6)	43	49.5 (SD 6.3)	<0.01[a]

[a] t-test

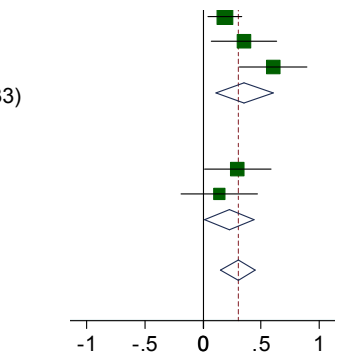
Synthesis with existing evidence-base

When the 2004 and post-2004 evidence-bases were collected together, there was an extremely heterogeneous collection of outcome measures for this domain. As a result, we have not been able to perform any quantitative synthesis of individual outcome measures on a natural scale.

4.6.1.2.2.1. Pooled multiple outcome measures

There were no new studies that measured functional outcomes at 24 weeks; therefore we pooled the functional outcome data from the studies in the previous assessment. This showed a significant benefit for donepezil at all doses compared to placebo, SMD=0.30 (95%CI 0.14, 0.45), p<0.001. See *Figure 12*. The data set used for this meta-analysis can be found in Appendix 6.

FIGURE 12 *Random-effects meta-analysis: functional outcomes (SMD) at 24wk – donepezil (all dosages) v. placebo*



Study	SMD	(95%CI)	Wght
LOCF analysis			
Burns et al. (1999)[97]	0.182	(0.036, 0.327)	32.1
Homma et al. (2000)[101]	0.352	(0.074, 0.630)	18.2
Gauthier et al. (2002)[98]	0.598	(0.309, 0.887)	17.4
subtotal (*Q*=6.66 [*p* on 2 d.f.=0.036]; I^2=70.0%; τ^2=0.033)	**0.353**	**(0.107, 0.599)**	**67.6**
		p=0.005	
OC population			
Mohs et al. (2001)[103]	0.293	(0.008, 0.578)	17.6
Winblad et al. (2001)[110]	0.136	(-0.192, 0.465)	14.7
subtotal (*Q*=0.5 [*p* on 1 d.f.=0.480]; I^2=0.0%; τ^2=0.000)	**0.226**	**(0.010, 0.441)**	**32.4**
		p=0.040	
Overall pooled estimate	**0.298**	**(0.144, 0.452)**	
(*Q*=7.36 [*p* on 4 d.f.=0.118]; I^2=45.6%; τ^2=0.014)		*p*<0.001	
Inter-stratum heterogeneity: *p*=0.656			
Small-study effects: Egger's *p*=0.363			

4.6.1.2.3. Behavioural and mood

In 2004 the assessment group reported that:

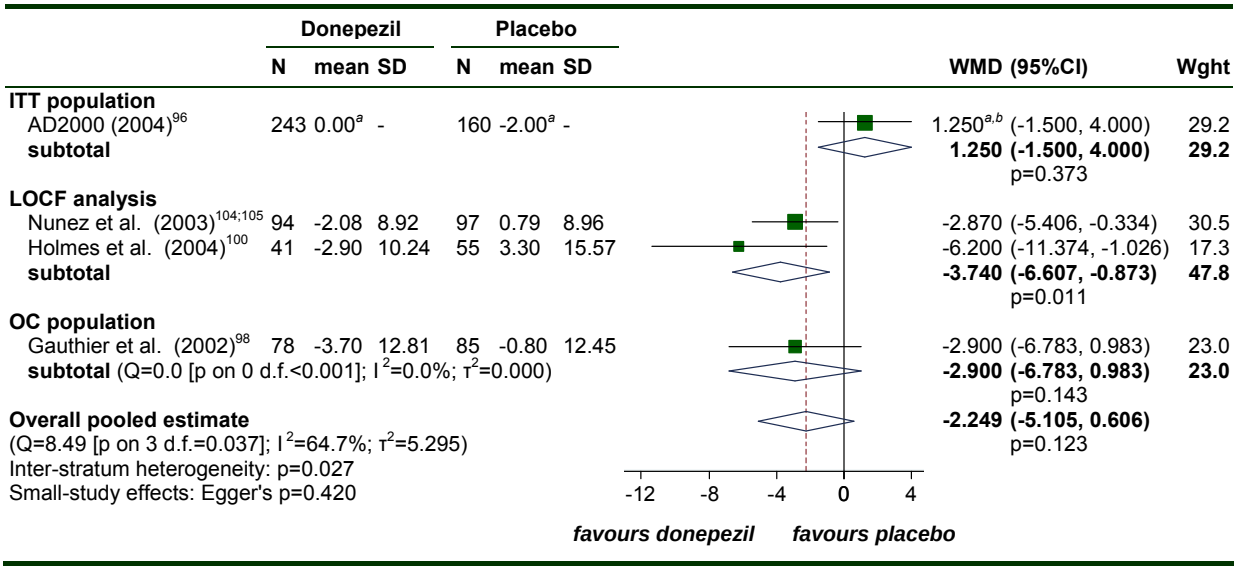
> *"The NPI was used as a measure of mood and behaviour in four RCTs. Data were varied but suggested that donepezil may have some effect in improving or limiting further deterioration on the NPI scale compared to placebo, at least over shorter durations of follow-up."*[2]

New data

None of the newly identified studies provided any additional data on the effect of donepezil as indicated by measures of behavioural function. Therefore we conducted random effects meta-analysis of the studies included in 2004 for the NPI at 12 and 24 weeks which showed no significant benefit from donepezil measured by the NPI, (see *Figure 13* and

Figure 14).

FIGURE 13 *Random-effects meta-analysis: NPI at 12wk (mean change from baseline) – donepezil (all dosages [all are 10mg/d])* v. *placebo*



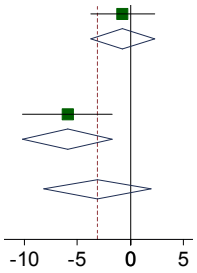
[a] score inverted from published figure to reflect usual direction of NPI

[b] WMD and error bars provided in publication; SE estimated on assumption that error-bars represent 95%CIs

FIGURE 14 *Random-effects meta-analysis: NPI at 24wk (mean change from baseline) – donepezil (all dosages [all are 10mg/d]) v. placebo*

	Donepezil			Placebo					
	N	mean	SD	N	mean	SD	WMD	(95%CI)	Wght
ITT population									
AD2000 (2004)[96]	209	-1.00[a]	-	225[a]	0.00	-	-0.750[a,b]	(-3.750, 2.250)	54.2
subtotal							**-0.750**	**(-3.750, 2.250)**	**54.2**
								p=0.624	
LOCF analysis									
Gauthier et al. (2002)[98]	97	-5.00	15.76	104	0.92	14.58	-5.920	(-10.126, -1.714)	45.8
subtotal							**-5.920**	**(-10.126, -1.714)**	**45.8**
								p=0.006	
Overall pooled estimate							**-3.116**	**(-8.165, 1.932)**	
								p=0.226	

(*Q*=3.85 [*p* on 1 d.f.=0.050]; I^2=74.0%; τ^2=9.891)
Inter-stratum heterogeneity: *p*=0.050
Small-study effects: not calculable



[a] score inverted from published figure to reflect usual of direction NPI

[b] WMD and error bars provided in publication; SE estimated on assumption that error-bars represent 95%CIs

4.6.1.2.3.1. Pooled multiple outcome measures

Because NPI is the only outcome measure used in this domain of the evidence-base, it was not necessary to pool outcomes on a standardised level.

4.6.1.2.4. Global effect

Loveman and colleagues summarised their findings on global outcomes comparing donepezil and placebo as:

> *"Seven RCTs assessed the effect of donepezil compared to placebo on the CGIC or CIBIC plus, showing overall that CGIC/CIBIC-plus scores were statistically significantly better with donepezil. The range of scores varied between the included studies. Higher proportions of participants receiving donepezil were considered as responders to treatment although this was not compared statistically in many cases. On the CDR scale trends were also demonstrated towards improved global function in the donepezil treated groups compared to the placebo groups in five trials but statistical significance was not demonstrated. In one unpublished trial with participants with mild AD, no benefit on the CDR was noted in the donepezil treated group."*[2]

New data

Only one of the new studies measured global outcomes.[115] They also found significant benefit on the clinical dementia rating scale (CDR), I=1.2 (SD 0.2), C=2.0 (SD 0.2), P<0.01, (see *Table 8*).

TABLE 8 *Measures of global effect in included studies – donepezil* v. *placebo*

Study	Subgroup	Outcome	Type	Donepezil		Placebo		*p*
				N	Mean	N	Mean	
Peng et al. (2005)[115]	OC population	Clinical Dementia Rating – 12wk	absolute value	46	1.2 (SD 0.2)	43	2 (SD 0.2)	<0.01[a]

[a] *t*-test

Synthesis with existing evidence-base

Clinician interview-based impression of change-plus

Only the previously included studies had data for meta-analysis of the CIBIC-plus. We pooled studies at 12 and 24 weeks and found that at both timepoints there was a significant overall pooled estimate of benefit from donepezil at 10 mg/d compared to placebo (12 weeks WMD=-0.38 (95%CI -0.49, -0.26), p<0.001 and 24 weeks WMD=-0.43 (95%CI-0.55, -0.31), p<0.001, (*Figure 15* and *Figure 16*). Meta-analyses of CIBIC-plus results for 5 mg/d and all doses combined can be found in Appendix 5.

FIGURE 15 *Random-effects meta-analysis: CIBIC-plus at 12wk (mean change from baseline) – donepezil (10mg/d) v. placebo*

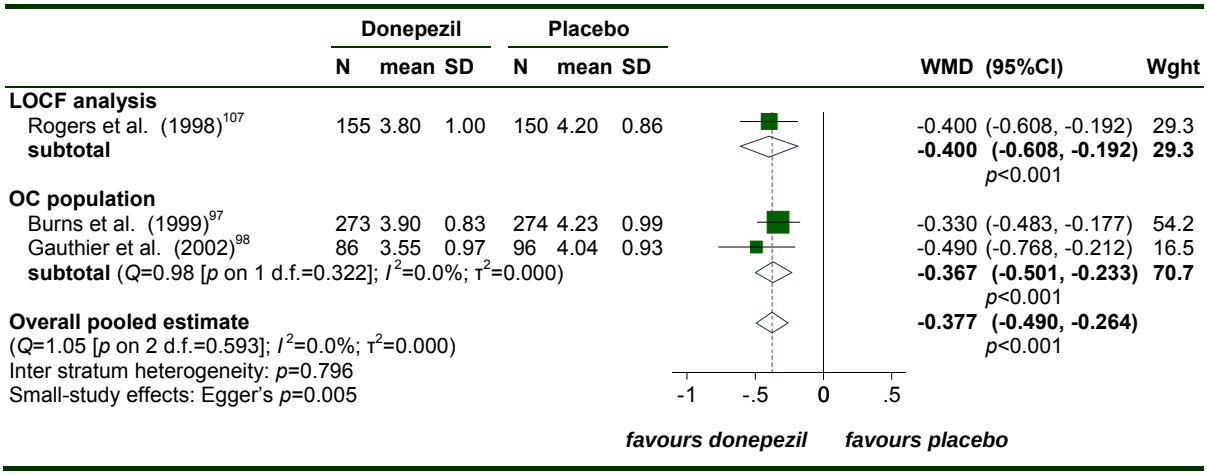
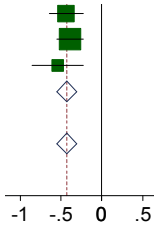


FIGURE 16 *Random-effects meta-analysis: CIBIC-plus at 24wk (mean change from baseline) – donepezil (10mg/d) v. placebo*

	Donepezil			Placebo				
	N	mean	SD	N	mean	SD	WMD (95%CI)	Wght
LOCF analysis								
Rogers et al. (1998)[106]	149	4.07	0.85	152	4.51	0.99	-0.440 (-0.648, -0.232)	33.2
Burns et al. (1999)[97]	273	4.13	0.99	274	4.52	0.99	-0.390 (-0.556, -0.224)	52.1
Gauthier et al. (2002)[98]	98	4.00	1.19	105	4.55	1.08	-0.545 (-0.858, -0.232)	14.7
subtotal (*Q*=0.75 [*p* on 2 d.f.=0.687]; I^2=0.0%; τ^2=0.000)							**-0.429 (-0.549, -0.309)**	**100.0**
							p<0.001	
Overall pooled estimate							**-0.429 (-0.549, -0.309)**	
(*Q*=0.75 [*p* on 2 d.f.=0.687]; I^2=0.0%; τ^2=0.000)							*p*<0.001	
Small-study effects: Egger's *p*=0.031								

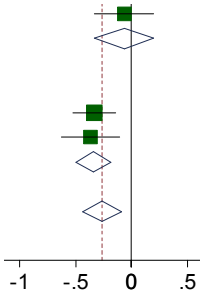


Clinical Dementia Rating

The pooled results on the CDR scale showed a significant advantage from taking donepezil at 12 and 24 weeks follow-up (12 weeks WMD=-0.26 (95%CI -0.44, -0.09), p<0.003 and 24 weeks WMD=-0.57 (95%CI-0.85, -0.29), p<0.001 (see *Figure 17* and *Figure 18*).

FIGURE 17 *Random-effects meta-analysis: Clinical dementia rating at 12wk (mean change from baseline) – donepezil (all dosages) v. placebo*

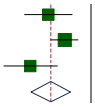
	Donepezil			Placebo				
	N	mean	SD	N	mean	SD	WMD (95%CI)	Wght
LOCF analysis								
Rogers et al. (1998)[107]	310[a]	-0.20	1.37	150	-0.14	1.35	-0.064 (-0.328, 0.200)	28.7
subtotal (*Q*=0.0 [*p* on 0 d.f.<0.001]; I^2=0.0%; τ^2=0.000)							**-0.064 (-0.328, 0.200)**	**28.7**
							p=0.633	
OC population								
Burns et al. (1999)[97]	544[a]	-0.18	1.32	274	0.15	1.32	-0.330 (-0.522, -0.138)	42.1
Homma et al. (2000)[101]	116	-0.11	0.94	112	0.25	1.06	-0.363 (-0.623, -0.102)	29.2
subtotal (Q=0.03 [p on 1 d.f.=0.852]; I^2=0.0%; τ^2=0.000)							**-0.341 (-0.496, -0.187)**	**71.3**
							p<0.001	
Overall pooled estimate							**-0.263 (-0.435, -0.091)**	
(Q=3.19 [p on 2 d.f.=0.203]; I^2=37.3%; τ^2=0.009)							*p*=0.003	
Inter stratum heterogeneity: p=0.076								
Small-study effects: Egger's *p*=0.664								



[a] pooled 5mg/d and 10mg/d arms

FIGURE 18 *Random-effects meta-analysis: Clinical dementia rating at 24wk (mean change from baseline) – donepezil (all dosages) v. placebo*

	Donepezil			Placebo				
	N	mean	SD	N	mean	SD	WMD (95%CI)	Wght
LOCF analysis								
Rogers et al. (1998)[106]	305[a]	-0.01	1.73	153	0.58	1.73	-0.595 (-0.931, -0.259)	30.2
Burns et al. (1999)[97]	544[a]	0.00	1.81	274	0.37	0.99	-0.370 (-0.563, -0.178)	42.6
Homma et al. (2000)[101]	116	-0.10	1.29	112	0.75	1.59	-0.850 (-1.226, -0.474)	27.2
subtotal (*Q*=5.37 [*p* on 2 d.f.=0.068]; I^2=62.7%; τ^2=0.038)							**-0.568 (-0.849, -0.288)**	**100.0**



		p<0.001
Overall pooled estimate		**-0.568 (-0.849, -0.288)**
(*Q*=5.37 [*p* on 2 d.f.=0.068]; I^2=62.7%; τ^2=0.038)		*p*<0.001
Small-study effects: Egger's *p*=0.182		

favours donepezil ***favours placebo***

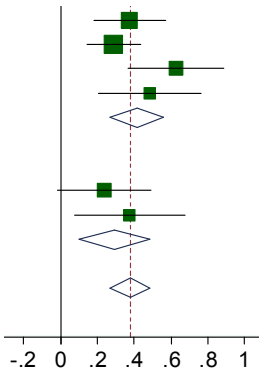
[a] pooled 5mg/d and 10mg/d arms

4.6.1.2.4.1. Pooled multiple outcome measures

We did not find any new studies that measured global outcomes at 24–26 weeks; therefore we pooled the global outcome data from the studies in the previous assessment. This showed a significant benefit for donepezil at all doses compared to placebo, SMD=0.38 (95%CI 0.27, 0.48), p<0.001. See *Figure 19*. The data set used in this meta-analysis can be found in Appendix 6.

FIGURE 19 *Random-effects meta-analysis: global outcomes (SMD) at 24–26wk – donepezil (all dosages) v. placebo*

Study	SMD	(95%CI)	Wght
LOCF analysis			
Rogers et al. (1998)[106]	0.375	(0.178, 0.571)	20.3
Burns et al. (1999)[97]	0.288	(0.142, 0.434)	29.4
Homma et al. (2000)[101]	0.626	(0.370, 0.883)	13.6
Gauthier et al. (2002)[98]	0.482	(0.202, 0.761)	11.9
subtotal (*Q*=5.56 [*p* on 3 d.f.=0.135]; I^2=46.0%; τ^2=0.010)	**0.414**	**(0.270, 0.557)**	**75.3**
		p<0.001	
OC population			
Winblad et al. (2001)[110]	0.236	(-0.017, 0.488)	14.0
Gauthier et al. (2002)[98]	0.375	(0.076, 0.673)	10.7
subtotal (*Q*=0.48 [*p* on 1 d.f.=0.486]; I^2=0.0%; τ^2=0.000)	**0.294**	**(0.101, 0.486)**	**24.7**
		p=0.003	
Overall pooled estimate (*Q*=6.74 [*p* on 5 d.f.=0.241]; I^2=25.8%; τ^2=0.005)	**0.377**	**(0.270, 0.484)**	
Inter-stratum heterogeneity: *p*=0.406		*p*<0.001	
Small-study effects: Egger's *p*=0.289			



favours placebo ***favours donepezil***

Meta-regression

We also conducted meta-regression analysis to explore the statistical heterogeneity across studies, looking at age, age and baseline MMSE (as a proxy for disease severity). Only one graph showed a significant relationship, this was between baseline MMSE and functional outcomes at 24 weeks for all doses of donepezil giving a meta-regression estimate of α=1.456; β=-0.065; *p*=0.009. However, due to the small number of studies in each analysis and that the data were assessed at a population level (which may not reflect the individual level) we felt that these results may be ambiguous and so have placed them in Appendix 7 in case they are of interest.

4.6.1.2.5. Quality of life

None of the included studies provided any additional data on QoL with donepezil compared with placebo. Accordingly, the data presented in the 2004 review (p. 32) represent a complete and current summary of randomised evidence on this subject.

4.6.1.2.6. Safety

None of the five newly identified studies provide data on AEs observed under randomised conditions.

Peng and colleagues[115] present limited safety data conflating their randomised study with data from a parallel observational study. Among the total of 145 individuals who took donepezil, 7 (4.8%) experienced an AE (it appears that one each experienced dizziness, nausea, inappetence, mild diarrhoea, constipation, fatigue, and agitation). Four of these seven cases stopped taking medicine. Among cases in the placebo group of the randomised trial, 2 (4.7%) experienced dizziness and stopped medication for this reason.

4.6.1.3. Summary: donepezil v. placebo

We found an additional five RCTs to add to the 14 previously found by Loveman and colleagues (2004), none of the new studies were of good quality or had a follow-up longer than six months.

Pooled cognitive outcomes showed a significant benefit from donepezil measured by the ADAS-cog and MMSE with greater benefit shown at 24 weeks (ADAS-cog: WMD=-2.90 (95%CI -3.61, -2.18), p<0.001, MMSE: WMD=1.21 (95%CI 0.84, 1.57), p<0.001). The pooled estimates of all cognitive outcomes likewise showed a benefit from donepezil at 24–26 weeks follow-up.

Only one new study looked at functional outcomes for this comparison; at 12 weeks this showed a significant gain for those taking donepezil (mean difference: I=40.5 (SD 7.6), C=49.5 (SD 6.3), p<0.01). At 24 weeks there was only data from the 2004 assessment trials, the results from all the studies reporting functional outcomes were pooled; this analysis again showed a significant benefit from taking donepezil.

None of the new studies measured behavioural outcomes; the pooled estimates from the previous assessment, using the NPI, failed to show a significant gain on behavioural outcomes at either 12 or 24 weeks.

Just one new study looked at global outcomes; this showed a benefit from taking donepezil on the CDR (I=1.2 (SD 0.2), C=2.0 (SD 0.2), P<0.01). The pooled results for the CIBIC-plus scale were only from the previous TAR, they showed a significant advantage from donepezil at 12 and 24 weeks follow-up (24 weeks: WMD=-0.43 (95%CI-0.55, -0.31), p<0.001. When both the global outcome measures were pooled at 24–26 weeks, the results again showed a significant benefit from donepezil.

None of the new studies provided additional data on QoL or safety under randomised conditions.

The new studies found have added to the body of evidence showing a benefit from donepezil compared to placebo for cognitive, functional and global outcomes. However, there is no new or pooled evidence to show a behavioural benefit from donepezil v. placebo in people with mild-to-moderate AD. All but two of the studies included in these meta-analyses calculated their missing data points using LOCF or OC methods, thereby potentially biasing their results in favour of donepezil.

4.6.1.4. Graphical summary of donepezil v. placebo

The first thing that is noticeable when comparing *Figure 20* and *Figure 21* with *Figure 22* is that no new large studies have been undertaken since 2004 and that only one study was multi-centre. Closer examination shows that the quality of trials has not improved and, with the exception of Peng and colleagues, studies only measured cognitive abilities. As in the previous review these outcomes showed that patients benefited from taking donepezil. While the new trials add to the precision of our knowledge of the effects of donepezil on cognitive measures in mixed mild/moderate AD populations, none of the new studies were in the mild AD population. This means that we are still dependent on one moderately sized RCT from the 2004 review that looked at cognitive outcomes in the mild AD population (Seltzer and colleagues 2004[109]), which showed a cognitive benefit for this group from taking 10 mg/day of donepezil.

FIGURE 20 *Summary of studies included in the 2004 review – donepezil v. placebo*

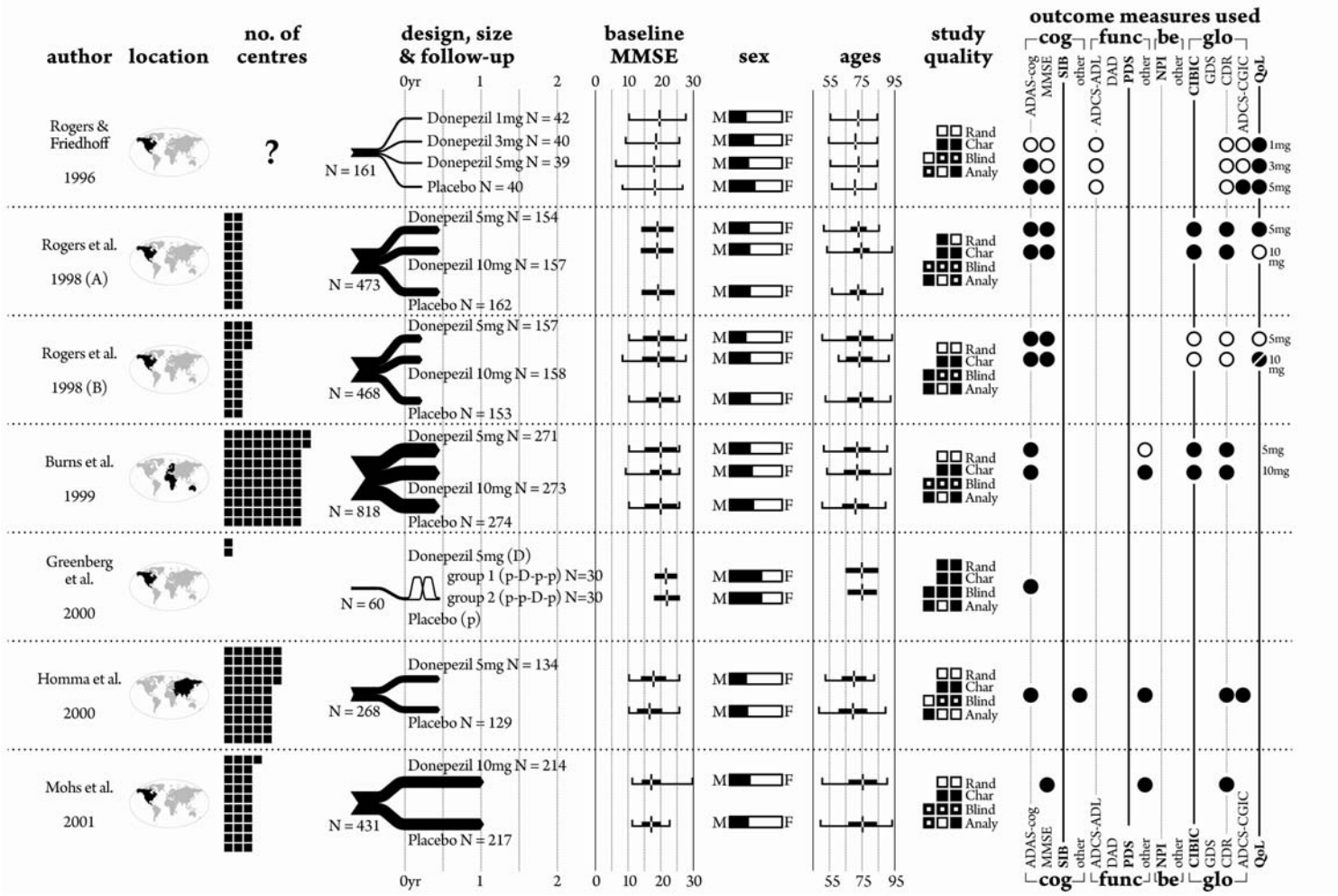


FIGURE 21 *Summary of studies included in the 2004 review – donepezil v. placebo cont.*

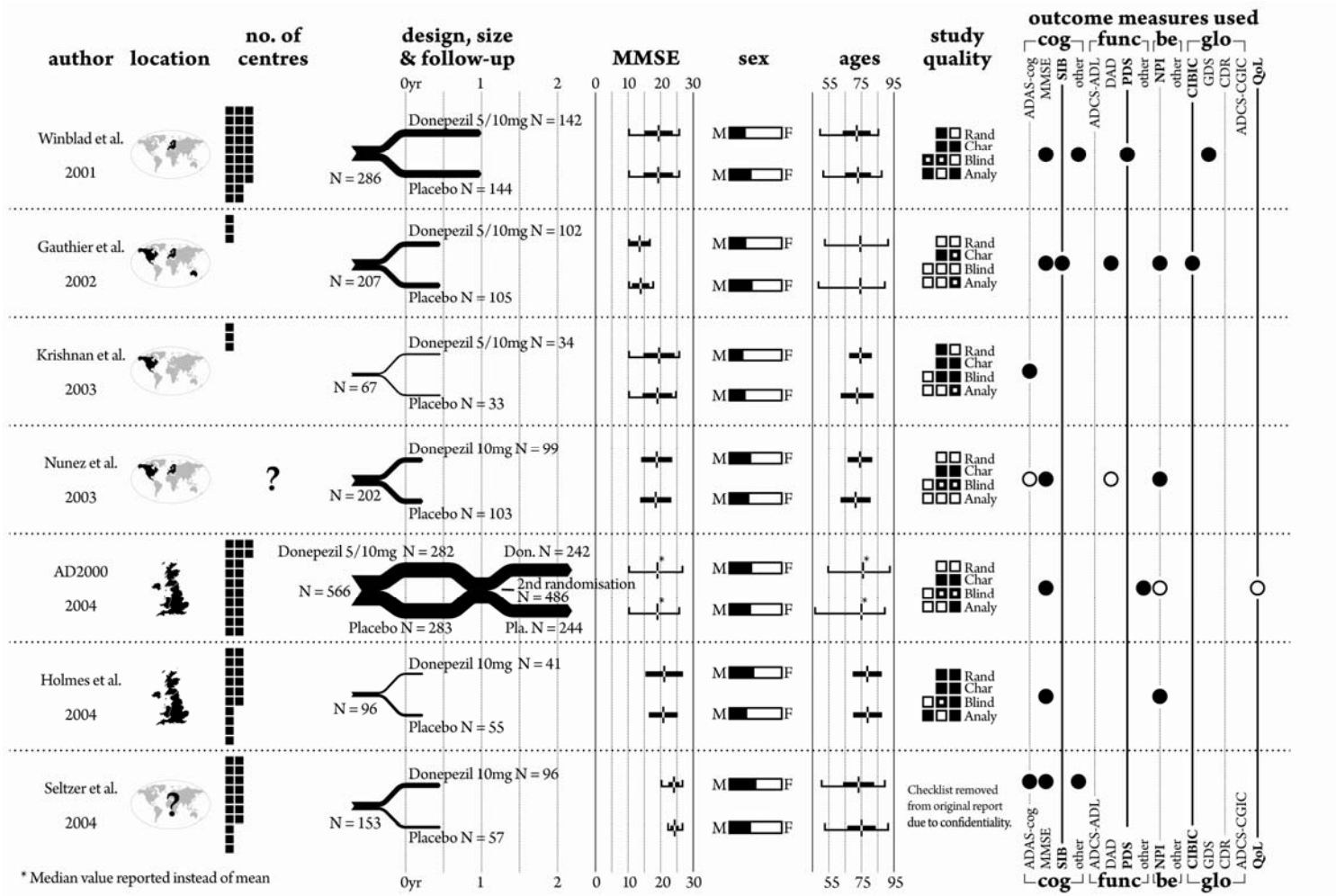
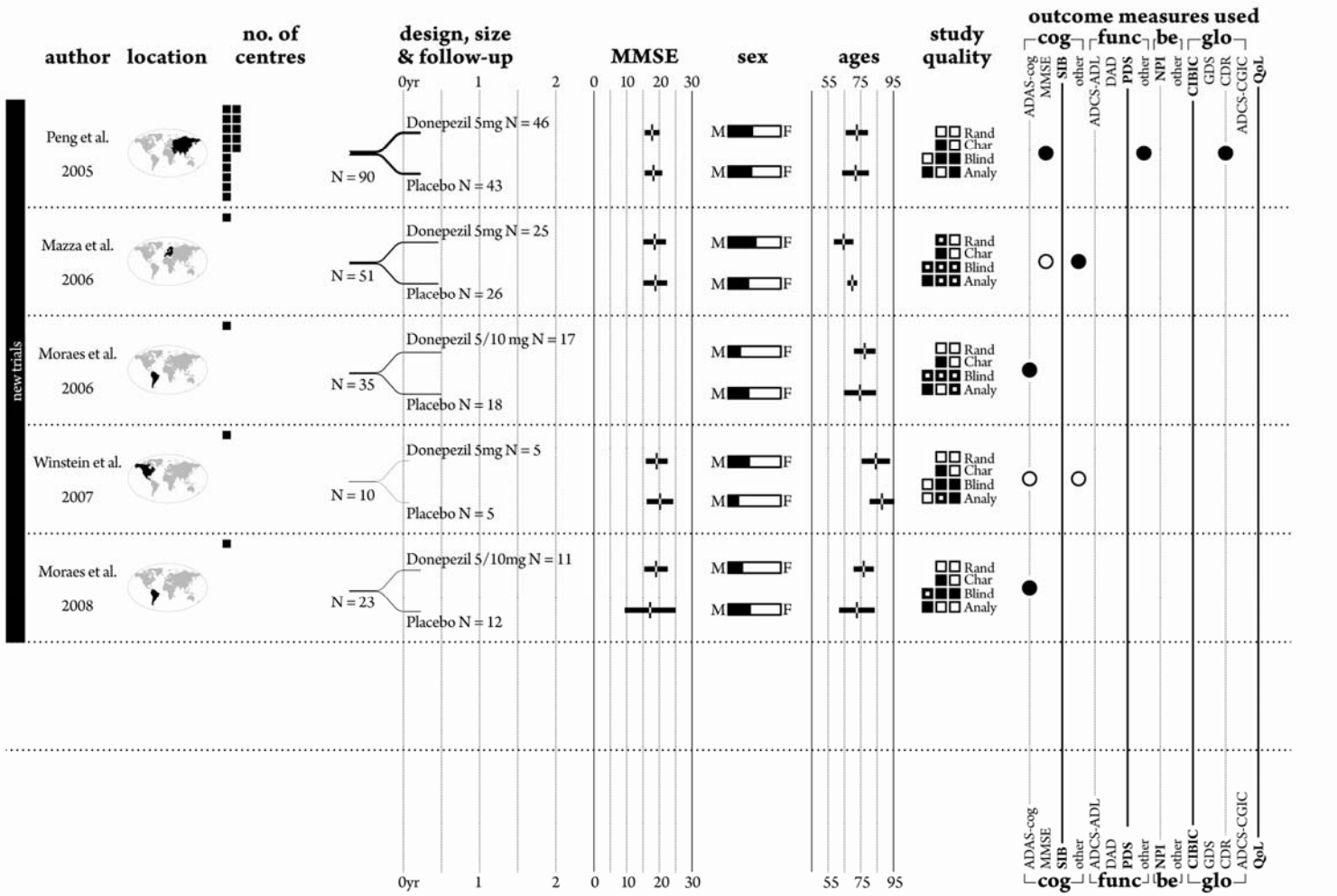


FIGURE 22 *Summary of new studies included in the 2010 review – donepezil v. placebo*



4.6.2. Galantamine *v.* placebo

4.6.2.1. Identified evidence

For the pairwise comparison between galantamine and placebo, the 2004 review included six RCTs, those reported by Raskind and colleagues. (2000),[117] Rockwood and colleagues. (2001),[118] Tariot and colleagues. (2000)[119] (supplemented by additional data from Cummings and colleagues. 2004[120]), Wilcock and colleagues. (2000),[121] Wilkinson & Murray (2001),[122] and Wilkinson et al. (2000),[123]. However, it is apparent that two of these publications – Rockwood and colleagues. (2001)[118] and Wilkinson and colleagues. (2000)[123] – report the same population. Accordingly, we have only entered Rockwood and colleague's primary publication[118] in syntheses, below.

We identified an additional three moderately good to poor quality RCTs, meeting few of the quality criteria indicated in section 4.1.4.

The primary publication of the GAL-INT-6 study, written by Erkinjuntti and colleagues in 2002,[124] was correctly excluded from the 2004 review because it conflated participants with multiple forms of dementia. However, we were able to include a subsequent paper – Bullock and colleagues 2004[94] – reporting the AD-specific subgroup of this trial. A summary of their design characteristics can be found in *Table 9* and the interventions, comparators and baseline characteristics of the participants in *Table 10*. A summary of the markers of internal validity is presented in *Table 11*.

TABLE 9 *Design of included studies – galantamine* v. *placebo*

Study details	Inclusion criteria	Exclusion criteria	Methodological notes	Other
Brodaty et al. (2005)[89] **Design:** Parallel double-blind RCT **Country:** United States, Australia, Canada, South Africa, and New Zealand **No. of centres:** 93 **No.**	mild to moderate probable AD (NINCDS-ADRDA) ADAS-cog/11 ≥18 history of cognitive decline that was gradual in onset and progressive over a period of ≥6mo	other neurodegenerative disorders or cognitive impairment due to acute cerebral trauma, hypoxic cerebral damage, vitamin deficiency states, infection, primary or metastatic cerebral neoplasia, significant endocrine or metabolic disease, or mental retardation vascular dementia or evidence of clinically active cerebrovascular disease history of epilepsy or convulsions; current clinically	**Sample attrition / dropout:** 768 of 971 completed study. 203 withdrew after allocation: did not receive treatment (n=6); adverse event (n=67); withdrew consent (n=62); noncompliance (n=29); lost to follow-up (n=10); insufficent response (n=10); death (n=5); other reasons (n=3). No differences between groups. **Randomisation and allocation:** Randomization to treatment was determined by calling an interactive voice	**Therapy common to all participants:** 1mo placebo run-in prior to treatment allocation **Study Funding:** none reported **Other conflicts:** Lead author declares consultancy fees, a grant, and sponsored speaking engagements from Janssen

Study details	Inclusion criteria	Exclusion criteria	Methodological notes	Other
randomised: 971 **Maximum follow-up:** 26 **MMSE range included:** 10–24	living with or regular daily visits from a responsible caregiver (≥5d/wk)	significant psychiatric disease; active peptic ulcer; clinically significant hepatic, renal, pulmonary, metabolic, or endocrine disturbances; clinically significant urinary outflow obstruction; clinically significant cardiovascular disease use of any agent for the treatment of dementia (approved, experimental, or over the counter) including, but not limited to, nootropic agents, cholinomimetic agents, estrogens taken without medical need, chronic nonsteroidal anti-infl ammatory agents or cycloxygenase-2 inhibitors (>30 consecutive days, regardless of indication), and vitamin E (unless a stable dose had been taken for ≥6mo prior to trial initiation).	response system. The subject number and treatment code (which corresponded to a specific medication kit) was randomly generated after the caller at the site provided the requested subject details. All treatments were supplied in opaque, size-0 gelatin capsules that were identical in appearance, taste and smell. All subjects received 1 capsule twice daily. **Power calculation:** Powered at >95% to detect a 2.5-point (SD 6.2) difference in ADAS-cog/11 score and at 90% to detect a 15% difference between active and placebo groups in their CIBIC-plus responder rates, assuming a 55% placebo responder rate (no change/improved CIBIC-plus score). Required sample size not explicitly reported.	
Bullock et al. (2004)[94] **Design:** Parallel double-blind RCT **Country:** 'Including' Canada, Denmark, Finland, France, Germany, Israel, The Netherlands, Poland, UK **No. of centres:** 62 **No. randomised:** 285 **Maximum follow-up:** 26 **MMSE range included:** 10–25	Probable vascular dementia (NINDS-AIREN definition) or AD + CVD (NINCDS-ADRDA definition) (with CVD evidenced by CT or MRI) Score >=12 on 11-item subscale of of AD assessment scale presence of focal neurological signs disease onset at between 40 and 90 years of age	neurogenerative disorders cognitive impairmentresulting from other cerebral trauma cerebral neoplasia mental retardation vitamin deficiency significant endocrine or metabolic disease clinically significant coexitsng medical conditions significant cardiovascular disease that would likely limit the patinet's ability to complete the study current use of agents for the treatment of dementia recent history (within 30 days) of treatment with other investigational agents history of alcohol or drug abuse	**Sample attrition / dropout:** 230 of 285 completed study **Randomisation and allocation:** Randomisation was conducted using a 'computer-generated code' (no further details provided). No details provided about appearance, taste, or smell of placebo. **Power calculation:** Not reported	**Therapy common to all participants:** 1mo single-blind placebo run-in prior to treatment allocation **Study Funding:** None reported **Other conflicts:** None reported **Notes:** Follow-up also at 32 and 52 weeks during the open-label phase of the trial Unable to calculate attrition n, as using percentages quoted in the text gives non-whole numbers
Rockwood et al. (2006)[125] **Design:** Parallel double-blind RCT **Country:**	Probable Alzheimer's disease (NINCDS-ADRDA criteria) ADAS-cof	Resident in nursing home Disabling communication difficulties (problems in language, speech, vision or hearing) Other active medical issues or	**Sample attrition / dropout:** 109 of 130 completed study. 21 withdrew after allocation: adverse event n=7; noncompliance n=6; insufficient response n=4; lost to follow-up n=1;	**Therapy common to all participants:** None reported **Study Funding:** Janssen-Ortho Canada (80%) and the Canadian Institutes of Health Research (20%) (grant no. DCT-49981).

Study details	Inclusion criteria	Exclusion criteria	Methodological notes	Other
Canada **No. of centres:** 10 **No. randomised:** 130 **Maximum follow-up:** 16 **MMSE range included:** 10–25	score >=18 Daily contact with a responsible caregiver	competing causes of dementia Patients who had taken anti-dementia medications within 30 days before screening for study enrolment Hypersensitivity to cholinomimetic agents or bromide Participation in other galantamine trials	withdrew consent n=2; died n=1. More patients in the galantamine group (n=5_ withdrew due to adverse events than in the placebo group (n=2), otherwise no difference between groups. **Randomisation and allocation:** Randomization was determined immediately before medication was administered by research nurse phoning into a contracted, interactive voice-response system for an assignment number. Nurse was blind to the number's meaning in terms of treatment assignment. Randomisation was in blocks of 2, by site, to decrease the chance of incomplete blocks (the GAS instrument was new to investigators at the study sites and that some sites might have had to withdraw if investigators did not know how to complete it) **Power calculation:** Authors state that on the basis that the GAS instrument can be more responsive than standard measures because it is personalized, this attribute had not been tested in a controlled trial in dementia. For the exploratory analysis, the sample size was estimated from the authors' limited experience with GAS in anti-dementia drug trials. Assuming a moderate effect size of about 0.524 and a 15% dropout at 4 months, it was determined that 152 subjects would be required to detect differences at the 5% significance level (2-tailed) with 80% power. Authors recognized that this might not result in statistically significant results for the secondary outcomes, which were used to compare with the primary outcomes and with results from other studies.	The sponsor provided all medications and matching placebos, conducted on-site monitoring and gathered and electronically coded the case report forms. All data are held by the principal investigator (Kenneth Rockwood), who initiated and supervised all analyses. Janssen-Ortho received the paper 45 days before submission to verify protocol details. At the authors' request, Janssen-Ortho statisticians answered questions about the use of the mixed effects model but had no other input in the analyses. **Other conflicts:** Lead author has undertaken consultancies and received honoraria from Janssen Ortho, the study's co-sponsor, and from Pfizer, Novartis and Merck, and was also lead author of an earlier galantamine study. Lead author owns no stock in pharmaceutical companies. Lead author is part owner of DementiaGuide, which is developing a Web site to aid in goal setting for people with dementia. Co-authors: CM has received research grants from Janssen Ortho, Pfizer, Lundbeck and Novartis, but has received no personal payments; MG has received honoraria and travel grants from Janssen Ortho, Pfizer and Merck; SF and XS have no conflicts of interest to declare. **Notes:** Five patients (2 in galantamine group, 3 in placebo group) had MMSE scores that were outside the 10-25 range stipulated in the inclusion criteria; 1 had an MMSE score <10, the other 4 had MMSE scores >25. Seven patients (4 in galantamine group, 3 in placebo group) had ADAS-Cog scores that were outside the >17 range stipulated in the inclusion criteria; in each case the score was below the lower limit, which indicated milder

Study details	Inclusion criteria	Exclusion criteria	Methodological notes	Other
				impairment

TABLE 10 *Interventions, comparators, and baseline characteristics of participants in included studies – galantamine* v. *placebo*

Study	Arm	Dose (mg/d)	Dosage details	N	Age	Sex (n male)	Race (n white)	Weight (kg)	Education (yrs)	Duration of dementia (mo)	ADAS-cog	MMSE
Brodaty et al. (2005)[89]	Galantamine PRC[a]	8–24	titrated from an initial dosage of 8mg/d for the first 4wk up to a maximum of 24mg/d in increments of 8 mg/day every 4wk after the placebo run-in whole dose given in single capsule in am; placebo given in pm	319	76.6 (SD 7.64)	114 (35.7%)	297 (93.1%)	68.6 (SD 14.2)[a]				18.0 (SD 3.97)
	Galantamine	8–24	titrated from an initial dosage of 8mg/d for the first 4wk up to a maximum of 24mg/d in increments of 8 mg/day every 4wk after the placebo run-in single capsules in am and pm	326	76.5 (SD 7.77)	118 (36.2%)	293 (89.9%)	68.3 (SD 15.9)				17.8 (SD 4.14)
	Placebo	-	single placebo dose in am and pm	320	76.3 (SD 8.03)	115 (35.9%)	289 (90.3%)	67.8 (SD 14.6)[b]				18.1 (SD 4.08)
Bullock et al. (2004)[94]	Galantamine	4–24	Titrated upwards in weekly 4mg increments over a period of 6wk, and then continued at this maintenance dose (24mg/day) for an additional 4.5mo	152	75.8 (SD 6.78)	73 (48.0%)		69.9 (SD 12.9)			22.7 (SD 9.37)[c]	20.5 (SD 3.95)
	Placebo	-	single placebo dose am and pm	86	77.6 (SD 6.12)	42 (48.8%)		67.0 (SD 13.0)			23.9 (SD 9.92)[d]	20.2 (SD 3.52)
Rockwood et al. (2006)[125]	Galantamine	8–24	Initial dose of 8mg/d (4mg twice daily) for 4 wk, followed by 16mg/d for another 4 wk. At the end of week 8, dose could be increased to 24mg/dy depending on tolerability. At week 12, patients were re-evaluated; the dose could then be reduced to 16mg/d if necessary, after which time it could not be changed.	64	77.0 (SD 8.00)	23 (35.9%)			11.0 (SD 3.00)		24.2 (SD 6.40)	20.8 (SD 3.30)
	Placebo	-	Sham titration schedule	66	78.0 (SD 8.00)	25 (37.9%)			11.0 (SD 3.00)		27.9 (SD 8.40)	19.9 (SD 4.20)

[a] *n*=318

[b] *n*=319

[c] *n*=148

[d] *n*=85

TABLE 11 *Markers of internal validity of included studies – galantamine* v. *placebo*

	Was the assignment to the treatment groups really random?	Was the treatment allocation concealed?	Were the groups similar at baseline in terms of prognostic factors?	Were the eligibility criteria specified?	Were outcome assessors blinded to the treatment allocation?	Was the care provider blinded?	Was the patient blinded?	Were the point estimates and measure of variability presented for the primary outcome measure?	Did the analyses include an intention-to-treat analysis?	Were withdrawals and dropouts completely described?
Brodaty *et al.* (2005)[89]	ADEQUATE	ADEQUATE	REPORTED – YES	ADEQUATE	ADEQUATE[a]	ADEQUATE	ADEQUATE	PARTIAL[b]	PARTIAL[c]	ADEQUATE
Bullock *et al.* (2004)[94]	PARTIAL[d]	UNKNOWN	REPORTED – YES	UNKNOWN	UNKNOWN	UNKNOWN	PARTIAL	ADEQUATE	PARTIAL[e]	ADEQUATE
Rockwood *et al.* (2006)[125]	ADEQUATE	ADEQUATE	REPORTED – NO[f]	INADEQUATE	PARTIAL	PARTIAL	PARTIAL	ADEQUATE	ADEQUATE	ADEQUATE

[a] treating healthcare providers + caregivers contributed to outcome assessment, though no reason to suspect blinding was compromised

[b] in one instance, data are repeated with different measures of dispersion

[c] LOCF analyses attempted; however, LOCF cohort is less than full sample size and decreases as follow-up extends

[d] Randomised using a computer-generated code (but not generated from a central office)

[e] ITT claimed, but n<original sample size

[f] Placebo group had more patients with moderate dementia

4.6.2.2. Evidence of clinical effectiveness

4.6.2.2.1. Cognition

The previous review summarised the evidence from cognition outcomes thus:

> *"Six published RCTs showed that galantamine appears to confer a statistically significant benefit to participants on the ADAS-cog scale when compared to placebo, whether reducing the deterioration or leading to some improvement in their condition. The benefit varies depending upon the dose of galantamine. The galantamine-placebo differences in ADAS-cog for 8mg/day was 1.3 points, 16mg/day 3.1 points, 18mg/day 1.7 points, 16 or 24mg/day 2.5 to 2.8 points; 24 to 32 mg/day 1.7 to 3.4 points and 36mg/day 2.3 points. The one unpublished RCT mirrored these positive effects of galantamine versus placebo. In addition, some 14-17% more of galantamine participants were classified as responders (improving by 4 or more points on the ADAS-Cog) than those on placebo."*[2]

New data

The results from the three new studies show that overall those treated with galantamine had improved scores on the ADAS-cog while those in the control group remained stable or declined. However, missing data were accounted for using LOCF and OC methods leading to a potential overestimate of the treatment effect.[80] *Table 12* summarises the cognitive results from the new studies.

TABLE 12 *Measures of cognition in included studies – galantamine* v. *placebo*

Study	Subgroup	Outcome	Type[a]	Arm	Galantamine		Placebo		p
					N	Mean	N	Mean	
Bullock et al. (2004)[94]	OC population	ADAS-cog – 6wk			148	-0.5 (SD 4.62)[b]	85	0.15 (SD 6.26)[b]	0.366[c]
		ADAS-cog – 13wk	MC		148	-1.48 (SD 4.32)[b]	85	0 (SD 6.03)[b]	0.031[c]
		ADAS-cog – 26wk	MC		147	-1.1 (SD 5.79)	83	2 (SD 5.56)	<0.001[c]
			A		147	21.5 (SD 10.5)	83	25.7 (SD 12)	0.006[c]
Brodaty et al. (2005)[89]	LOCF analysis	ADAS-cog – 8wk	MC	od[d]	287	-1.5 (SD 5.08)	293	0 (SD 5.14)	-
				bd[e]	294	-1.7 (SD 4.97)			-
		ADAS-cog – 12wk	MC	od[d]	290	-2[f] (SD 5.28)	296	0.2 (SD 5.33)	-
				bd[e]	296	-2.5 (SD 5.16)			-
		ADAS-cog – 26wk	MC	od[d]	291	-1.3 (SD 5.29)	296	1.2 (SD 5.68)	<0.001[g]
				bd[e]	296	-1.6 (SD 6.19)			<0.001[g]
	OC population	ADAS-cog – 8wk	MC	od[d]	284	-1.5 (SD 5.06)	289	0 (SD 5.1)	-
				bd[e]	286	-1.7 (SD 5.07)			-
		ADAS-cog – 12wk	MC	od[d]	269	-2.2 (SD 5.25)	275	0 (SD 5.14)	-
				bd[e]	268	-2.6 (SD 5.07)			-
		ADAS-cog – 26wk	MC	od[d]	240	-1.4 (SD 5.27)	248	1.3 (SD 5.67)	<0.001[g]
				bd[e]	227	-1.8 (SD 6.33)			<0.01[g]
Rockwood et al. (2006)[125]	LOCF analysis	ADAS-cog – 8wk	MC		62	-1.85 (SD 4.18)	65	-0.25 (SD 4.97)	-
		ADAS-cog – 16wk	MC		62	-1.6 (SD 5.38)	65	0.325 (SD 5.49)	-

[a] MC=mean change; A=absolute value

[b] estimated from figure

[c] student's t-test (calculated by reviewer)

[d] galantamine prolonged release once a day

[e] galantamine twice a day

[f] value given as +2 in paper; assumed to be a typographical error, as all other observations are negative, and corroborated by OC data, which are otherwise relatively consistent with LOCF figures

[g] ANOVA with factors for treatment and pooled country (United States v.ex-United States)

Synthesis with existing evidence-base

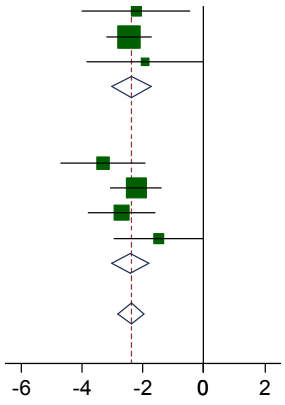
We pooled data from the new studies with those of the 2004 assessment by random-effects meta-analysis using the ADAS-cog as the outcome measure for ≤24mg/d at 12–16 weeks and 21–26 weeks post randomisation, no studies reported the MMSE. We also meta-analysed the data by >12 mg/d and all doses combined; these results can be found in Appendix 5.

ADAS-cog

Two new studies were included in the meta-analysis. The overall pooled estimates showed a benefit from galantamine compared to placebo at 12–16 and 21–26 weeks, 12–16 weeks WMD=-2.39 (95%CI -2.80, -1.97), p<0.001 and 21–26 weeks WMD=-2.96 (95%CI -3.41, -2.51), p<0.001 (see *Figure 23* and *Figure 24*).

FIGURE 23 *Random-effects meta-analysis: ADAS-cog at 12–16wk (mean change from baseline) – galantamine (maximum dose ≤24mg/d)* v. *placebo*

	Galantamine			Placebo				
	N	mean	SD	N	mean	SD	WMD (95%CI)	Wght
LOCF analysis								
Wilkinson & Murray (2001)[122]	136[a]	-0.63	6.45	82	1.60	6.34	-2.226 (-3.975, -0.477)	5.7
Brodaty et al. (2005)[89]	586[b]	-2.25	5.22	296	0.20	5.33	-2.453 (-3.192, -1.713)	31.9
Rockwood et al. (2006)[125]	62	-1.60	5.38	65	0.33	5.49	-1.925 (-3.816, -0.034)	4.9
subtotal (*Q*=4.61 [*p* on 2 d.f.=0.100]; I^2=56.6%; τ^2=0.689)							**-2.361 (-3.002, -1.720)**	**42.4**
							p<0.001	
OC population								
Raskind et al. (2000)[117]	131	-3.30	6.01	157	0.00	5.95	-3.300 (-4.688, -1.912)	9.1
Tariot et al. (2000)[119]	520[c]	-1.62	5.16	225	0.60	5.25	-2.225 (-3.042, -1.408)	26.1
Wilcock et al. (2000)[121]	156	-2.10	5.00	171	0.60	5.23	-2.700 (-3.809, -1.591)	14.2
Bullock et al. (2004)[94]	148	-1.48	4.32	85	0.00	6.03	-1.475 (-2.933, -0.017)	8.2
subtotal (*Q*=3.62 [*p* on 3 d.f.=0.306]; I^2=17.1%; τ^2=0.071)							**-2.414 (-3.034, -1.794)**	**57.6**
							p<0.001	
Overall pooled estimate							**-2.386 (-2.804, -1.969)**	
(*Q*=3.92 [*p* on 6 d.f.=0.688]; I^2=0.0%; τ^2=0.000)							*p*<0.001	
Inter stratum heterogeneity: *p*=0.921								
Small-study effects: Egger's *p*=0.759								



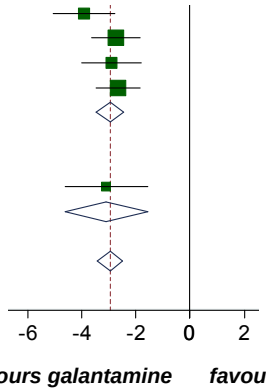
[a] 18mg/d and 24mg/d arms pooled

[b] once-daily prolonged release formulation and twice-daily standard formulation pooled

[c] 8mg/d, 16mg/d, and 24mg/d arms pooled

FIGURE 24 *Random-effects meta-analysis: ADAS-cog at 21–26wk (mean change from baseline) – galantamine (maximum dose ≤24mg/d) v. placebo*

	Galantamine			Placebo				
	N	mean	SD	N	mean	SD	WMD (95%CI)	Wght
LOCF analysis								
Raskind et al. (2000)[117]	202	-1.90	5.12	207	2.00	6.47	-3.900 (-5.029, -2.771)	16.1
Tariot et al. (2000)[119]	632[a]	-1.04	5.88	255	1.70	6.23	-2.741 (-3.633, -1.850)	25.8
Wilcock et al. (2000)[121]	220	-0.50	5.64	215	2.40	6.01	-2.900 (-3.996, -1.804)	17.1
Brodaty et al. (2005)[89]	587[b]	-1.45	5.76	296	1.20	5.68	-2.651 (-3.449, -1.854)	32.2
subtotal (*Q*=3.47 [*p* on 3 d.f.=0.324]; I^2=13.7%; τ^2=0.038)							**-2.957 (-3.471, -2.443)**	**91.1**
							p<0.001	
OC population								
Bullock et al. (2004)[94]	147	-1.10	5.79	83	2.00	5.56	-3.100 (-4.620, -1.580)	8.9
subtotal							**-3.100 (-4.620, -1.580)**	**8.9**
							p<0.001	
Overall pooled estimate							**-2.957 (-3.410, -2.505)**	
(*Q*=3.51 [*p* on 4 d.f.=0.476]; I^2=0.0%; τ^2=0.000)							*p*<0.001	
Inter-stratum heterogeneity: *p*=0.847								
Small-study effects: Egger's *p*=0.324								



[a] 8mg/d, 16mg/d, and 24mg/d arms pooled

[b] once-daily prolonged release formulation and twice-daily standard formulation pooled

4.6.2.2.1.1. Pooled multiple outcome measures

Because ADAS-cog is the only outcome measure used to assess cognitive effect in placebo-controlled trials of galantamine, it was not necessary to pool outcomes on a standardised level for this domain.

4.6.2.2.2. Functional

In 2004 the assessment group reported for functional outcomes that:

> *"Three RCTs assessed mean changes from baseline on the DAD scale, all reporting statistically significantly slower deterioration for those receiving galantamine 24-32mg/day compared to placebo. Two RCTs found that participants receiving 16mg/day and/or 24mg/day galantamine experienced a statistically significantly smaller deterioration on the ADCS/ADL compared to placebo."*[2]

New data

All three new RCTs measured functional outcomes. They found functional abilities had improved significantly more in the treatment group. *Table 13* summarises the results.

TABLE 13 *Measures of functional effect in included studies – galantamine* v. *placebo*

Study	Subgroup	Outcome	Type[a]	Arm	Galantamine		Placebo		p
					N	Mean / n (%)	N	Mean / n (%)	
Bullock et al. (2004)[94]	LOCF analysis	DAD – 26wk	MC		188	-1 (SD 15.8)[b]	97	-6 (SD 14.5)[b]	<0.01[c]
Brodaty et al. (2005)[89]	LOCF analysis	ADCS-ADL – 26wk	MC	od[d]	245[e]	0 (SD 7.51)	258[e]	-2.7 (SD 0.899)	<0.001[f]
			MC	bd[g]	242[e]	-1 (SD 0.778)			0.018[f]
	OC population	ADCS-ADL – 8wk	MC	od[d]	280	0.8 (SD 6.86)	294	-0.7 (SD 7.72)	-
			MC	bd[g]	292	0.9 (SD 7.18)			-
		ADCS-ADL – 12wk	MC	od[d]	276	0.4 (SD 6.65)	281	-0.3 (SD 7.71)	-
			MC	bd[g]	279	1.1 (SD 7.85)			-
		ADCS-ADL – 26wk	MC	od[d]	245	0 (SD 8.61)	258	-2.4 (SD 9.64)	0.003[f]
			MC	bd[g]	242	-1 (SD 8.87)[h]			0.088[f]
Rockwood et al. (2006)[125]	LOCF analysis	GAS (CR[i]) – 8wk	A		61	52.5 (SD 9.12)	66	52.2 (SD 6.97)	-
		GAS (CR[i]) – 16wk	A		61	54.8 (SD 9.36)	66	50.9 (SD 9.74)	0.02[k]
		GAS (PCR[j]) – 8wk	A		61	54.6 (SD 7.97)	66	52.5 (SD 8.57)	-
		GAS (PCR[j]) – 16wk	A		61	54.2 (SD 10.8)	66	52.3 (SD 9.12)	0.27[k]
	OC population[l]	GAS-VR: improved – 16wk	D		20	14 (70.0%)	30	8 (26.7%)	<0.01[m]
		GAS-VR: worsened – 16wk	D		20	2 (10.0%)	30	10 (33.3%)	-
		GAS VR: no change – 16wk	D		20	4 (20.0%)	30	12 (40.0%)	-

[a] MC=mean change; A=absolute value; D=dichotomous

[b] data extracted from publication using IPD in pooled analysis (Feldman et al. 2005 [126])

[c] test not specified

[d] galantamine prolonged release once a day

[e] sample size not provided (must be ≥ number of 26wk observed cases); precision likely to be underestimated

[f] ANOVA with factors for treatment and pooled country (United States v.ex-United States)

[g] galantamine twice a day

[h] different values for SE given in Table 2 (1.12) and Figure 4 (0.57) of publication; latter used as closer to range of dispersion reported in other arms

[i] clinician-rated

[j] patient-caregiver-rated

[k] ANOVA

[l] data extracted from secondary publication reporting subgroup with verbal repetition goals[127]

[m] mixed effects model, with dementia severity and treatment assignment as fixed effects, and the patient as the random effect

Synthesis with existing evidence-base

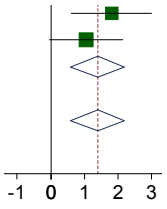
The data from the new trials were pooled, by random-effects meta-analysis, with those of the studies found in 2004. The outcome measures considered were the Alzheimer's Disease Cooperative Study-Activities of Daily Living (ADCS-ADL) and the Disability Assessment of Dementia (DAD).

ADCS-ADL

Results from the ADCS-ADL were pooled at 12–13 weeks follow-up and 21–26 weeks follow-up. The overall pooled estimates showed functional benefit from galantamine compared to placebo, 12–13 weeks WMD=1.39 (95%CI 0.59, 2.20), p<0.001 and 21–26 weeks WMD=2.23 (95%CI 1.33, 3.14), p<0.001, (*Figure 25* and *Figure 26*).

FIGURE 25 *Random-effects meta-analysis: ADCS-ADL at 12–13wk (mean change from baseline) – galantamine (≤24mg/d]) v. placebo*

	Galantamine			Placebo				
	N	mean	SD	N	mean	SD	WMD (95%CI)	Wght
OC population								
Tariot et al. (2000)[119]	530[a]	-0.44	8.07	235	-2.25	7.66	1.810 (0.613, 3.007)	45.2
Brodaty et al. (2005)[89]	555[b]	0.75	7.28	281	-0.30	7.71	1.052 (-0.034, 2.138)	54.8
subtotal (Q=0.85 [p on 1 d.f.=0.358]; I^2=0.0%; τ^2=0.000)							**1.394 (0.590, 2.198)**	**100.0**
							p<0.001	
Overall pooled estimate							**1.394 (0.590, 2.198)**	
(Q=0.85 [p on 1 d.f.=0.358]; I^2=0.0%; τ^2=0.000)							p<0.001	
Small-study effects: not calculable								

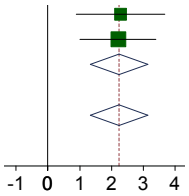


[a] 8mg/d, 16mg/d, and 24mg/d arms pooled

[b] once-daily prolonged release formulation and twice-daily standard formulation pooled

FIGURE 26 *Random-effects meta-analysis: ADCS-ADL at 21–26wk (mean change from baseline) – galantamine (≤24mg/d]) v. placebo*

	Galantamine			Placebo				
	N	mean	SD	N	mean	SD	WMD (95%CI)	Wght
LOCF analysis								
Tariot et al. (2000)[119]	637[a]	-1.52	9.47	262	-3.80	9.71	2.276 (0.889, 3.663)	42.7
Brodaty et al. (2005)[89]	487[b]	-0.50	5.36	258	-2.70	8.99	2.203 (1.007, 3.399)	57.3
subtotal (Q=0.01 [p on 1 d.f.=0.923]; I^2=0.0%; τ^2=0.000)							**2.234 (1.328, 3.140)**	**100.0**
							p<0.001	
Overall pooled estimate							**2.234 (1.328, 3.140)**	
(Q=0.01 [p on 1 d.f.=0.938]; I^2=0.0%; τ^2=0.000)							p<0.001	
Small-study effects: not calculable								



[a] 8mg/d, 16mg/d, and 24mg/d arms pooled

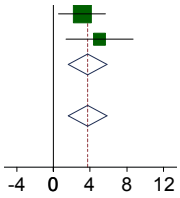
[b] once-daily prolonged release formulation and twice-daily standard formulation pooled

Disability Assessment for Dementia

The results from the DAD were pooled at 21–26 weeks follow-up. They again showed a significant benefit from galantamine compared with placebo, WMD=3.76 (95%CI 1.66, 5.86), p<0.001 (*Figure 27*).

FIGURE 27 *Random-effects meta-analysis: DAD at 21–26wk (mean change from baseline) – galantamine (≤24mg/d]) v. placebo*

	Galantamine			Placebo				
	N	mean	SD	N	mean	SD	WMD (95%CI)	Wght
LOCF analysis								
Wilcock et al. (2000)[121]	426[a]	-2.85	15.26	210	-6.00	15.65	3.152 (0.586, 5.717)	67.0
Bullock et al. (2004)[94]	188	-1.00	15.77	97	-6.00	14.48	5.000 (1.342, 8.658)	33.0
subtotal (Q=0.66 [p on 1 d.f.=0.417]; I^2=0.0%; τ^2=0.000)							**3.761 (1.661, 5.861)**	**100.0**
							p<0.001	
Overall pooled estimate							**3.761 (1.661, 5.861)**	
(Q=0.66 [p on 1 d.f.=0.417]; I^2=0.0%; τ^2=0.000)							p<0.001	
Small-study effects: not calculable								

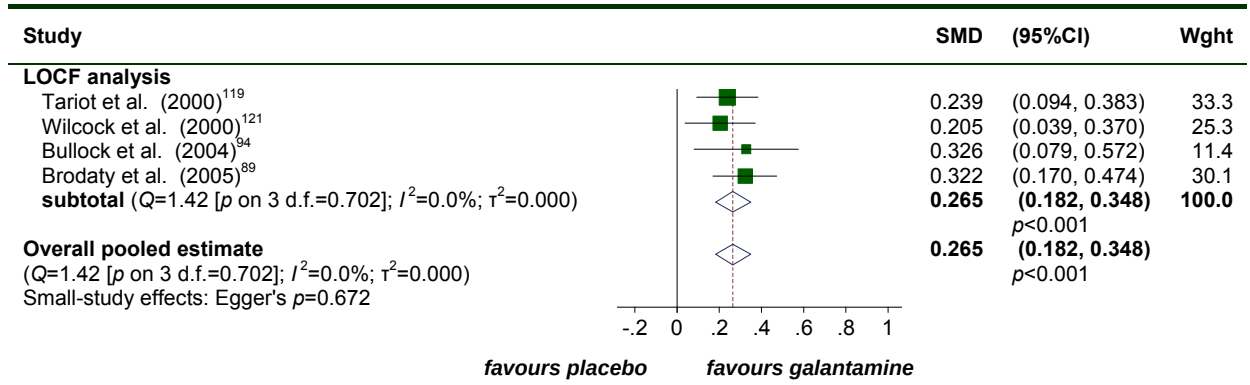


[a] 24mg/d and 32mg/d arms pooled

4.6.2.2.2.1. Pooled multiple outcome measures

Two new studies were added to the meta-analysis of combined functional outcome measures at 21–26 weeks. The overall pooled estimate showed a significant functional benefit from galantamine compared to placebo, SMD=0.27 (95%CI 0.18, 0.34), p<0.001 (*Figure 28*). The data set used in this meta-analysis can be found in Appendix 6

FIGURE 28 *Random-effects meta-analysis: functional outcomes (SMD) at 21–26wk – galantamine (all dosages) v. placebo*



With only four datapoints in this evidence-base, it would not be informative to perform meta-regression.

4.6.2.2.3. Behavioural and mood

In 2004 Loveman and colleagues summarised the behavioural results from this comparison as:

> *"Two published and one unpublished RCTs found that galantamine had some effect in improving or limiting further deterioration on the NPI scale compared to placebo. Differences in the mean change from baseline were statistically significant for doses of 16mg/day or over in one of the three studies."*[2]

New data

Only one included study – Brodaty and colleagues (2005)[89] – provided additional data on the effectiveness of galantamine in relieving behavioural symptoms of AD, when compared with placebo. However, this failed to show any statistically significant benefit. Data are shown in *Table 14*.

TABLE 14 *Measures of behavioural effect and mood in included studies – galantamine* v. *placebo*

Study	Subgroup	Outcome	Type	Arm	Galantamine		Placebo		p
					N	Mean / n (%)	N	Mean / n (%)	
Brodaty et al. (2005)[89]	LOCF analysis	NPI – 26wk	Mean change from baseline	od[a]	245[b]	-0.6 (SD 10.3)	258[b]	0.6 (SD 9.96)	0.941[c]
				bd[d]	242[b]	-0.9 (SD 11.4)			0.102[c]
	OC population	NPI – 26wk	Mean change from baseline	od[a]	245	-0.6 (SD 10.8)	258	0.1 (SD 13.2)	0.451[c]
				bd[d]	242	-1.2 (SD 12.9)			0.203[c]

[a] galantamine prolonged release formulation once a day

[b] sample size not provided (must be ≥ number of 26wk observed cases); precision likely to be underestimated

[c] ANOVA with factors for treatment and pooled country (United States *v.* ex-United States)

[d] galantamine standard formulation twice a day

Synthesis with existing evidence-base

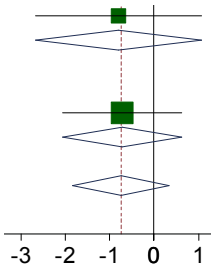
Neuropsychiatric Inventory

Only one new study added evidence to this meta-analysis. We looked for estimates of effectiveness at 13 and 21–26 weeks; at 13 weeks no significant benefit was found. However, at 21–26 weeks the overall pooled estimate favoured galantamine, WMD=-1.46 (95%CI -2.59, -0.34), p=0.012 (see *Figure 29* and *Figure 30*).

FIGURE 29 *Random-effects meta-analysis: NPI at 13wk (mean change from baseline) – galantamine (all dosages) v. placebo*

	Galantamine			Placebo				
	N	mean	SD	N	mean	SD	WMD (95%CI)	Wght
LOCF analysis								
Rockwood et al. (2001)[118]	241	-0.30	10.87	123	0.50	7.21	-0.800 (-2.672, 1.072)	33.8
subtotal							**-0.800 (-2.672, 1.072)**	**33.8**
							p=0.402	
OC population								
Tariot et al. (2000)[119]	529[a]	0.23	9.29	234	0.95	8.41	-0.719 (-2.056, 0.618)	66.2
subtotal							**-0.719 (-2.056, 0.618)**	**66.2**
							p=0.292	
Overall pooled estimate							**-0.746 (-1.835, 0.342)**	
							p=0.179	

(*Q*=0.0 [*p* on 1 d.f.=0.945]; I^2=0.0%; τ^2=0.000)
Inter-stratum heterogeneity: *p*=0.945
Small-study effects: not calculable



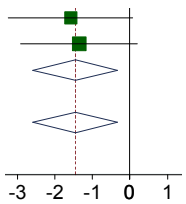
favours galantamine *favours placebo*

[a] 8mg/d, 16mg/d, and 24mg/d arms pooled

FIGURE 30 *Random-effects meta-analysis: NPI at 21–26wk (mean change from baseline) – galantamine (all dosages) v. placebo*

	Galantamine			Placebo				
	N	mean	SD	N	mean	SD	WMD (95%CI)	Wght
LOCF analysis								
Tariot et al. (2000)[119]	637[a]	0.43	11.85	262	2.00	11.33	-1.574 (-3.226, 0.078)	46.8
Brodaty et al. (2005)[89]	487[b]	-0.75	10.85	258	0.60	9.96	-1.349 (-2.900, 0.202)	53.2
subtotal (*Q*=0.04 [*p* on 1 d.f.=0.846]; I^2=0.0%; τ^2=0.000)							**-1.455 (-2.585, -0.324)**	**100.0**
							p=0.012	
Overall pooled estimate							**-1.455 (-2.585, -0.324)**	
(*Q*=0.04 [*p* on 1 d.f.=0.846]; I^2=0.0%; τ^2=0.000)							*p*=0.012	

Small-study effects: not calculable



favours galantamine *favours placebo*

[a] 8mg/d, 16mg/d, and 24mg/d arms pooled

[b] once daily prolonged release formulation and twice daily standard formulation pooled

4.6.2.2.3.1. Pooled multiple outcome measures

Because the NPI is the only outcome measure used in this domain of the evidence-base, it was not necessary to pool outcomes on a standardised level.

4.6.2.2.4. Global effect

The previous assessment in 2004 summarised the effectiveness of galantamine on global outcomes:

> *"Five published and one unpublished RCT assessed the effect of galantamine compared to placebo on the CIBIC plus, individually showing that higher proportions of participants receiving galantamine experience improvement in their condition compared to those on placebo (0% to 6.5% more participants). In contrast, a higher proportion of placebo participants tend to deteriorate (4% to 18% more participants). Also, a higher proportion of galantamine compared to placebo participants were considered to be responders to treatment with differences of between 4% (8mg/day) and 17% (24mg/day). When studies are pooled there are no statistically significant effects demonstrated."*[2]

New data

Two new studies were found that measured global outcomes.[89;125] Rockwood and colleagues found a benefit from galantamine measured by the CIBIC-plus compared to placebo at 13–16 weeks, (see *Table 15*).

TABLE 15 *Measures of global effect in included studies – galantamine* v. *placebo*

Study	Subgroup	Outcome	Type[a]	Arm	Galantamine		Placebo		p
					N	Mean / n (%)	N	Mean / n (%)	
Brodaty et al. (2005)[89]	LOCF analysis	CIBIC-plus score – 26wk	A	od[b]	291	4.21 (SD 1.1)	301	4.35 (SD 1.14)	NS[c]
				bd[d]	302	4.21 (SD 1.07)			NS[c]
		CIBIC-plus: markedly improved – 26wk	D	od[b]	291	3 (1.0%)	301	3 (1.0%)	0.712[e]
				bd[d]	302	3 (1.0%)			0.685[e]
		CIBIC-plus: moderately improved – 26wk	D	od[b]	291	14 (4.8%)	301	11 (3.7%)	0.621[e]
				bd[d]	302	15 (5.0%)			0.553[e]
		CIBIC-plus: minimally improved – 26wk	D	od[b]	291	49 (16.8%)	301	48 (15.9%)	0.856[e]
				bd[d]	302	46 (15.2%)			0.897[e]
		CIBIC-plus: no change – 26wk	D	od[b]	291	114 (39.2%)	301	111 (36.9%)	0.623[e]
				bd[d]	302	127 (42.1%)			0.224[e]
		CIBIC-plus: minimally worse – 26wk	D	od[b]	291	81 (27.8%)	301	80 (26.6%)	0.802[e]
				bd[d]	302	78 (25.8%)			0.907[e]
		CIBIC-plus: moderately worse – 26wk	D	od[b]	291	24 (8.2%)	301	41 (13.6%)	0.050[e]
				bd[d]	302	30 (9.9%)			0.201[e]
		CIBIC-plus: markedly worse – 26wk	D	od[b]	291	6 (2.1%)	301	7 (2.3%)	0.951[e]
				bd[d]	302	3 (1.0%)			0.336[e]
	OC population	CIBIC-plus score – 26wk	A	od[b]	246	4.19 (SD 1.13)	259	4.36 (SD 1.15)	NS[c]
				bd[d]	240	4.21 (SD 1.11)			NS[c]
		CIBIC-plus: markedly improved – 26wk	D	od[b]	246	3 (1.2%)	259	3 (1.2%)	0.728[e]
				bd[d]	240	3 (1.3%)			0.751[e]
		CIBIC-plus: moderately improved – 26wk	D	od[b]	246	14 (5.7%)	259	9 (3.5%)	0.327[e]
				bd[d]	240	24 (5.8%)			0.298[e]
		CIBIC-plus: minimally improved – 26wk	D	od[b]	246	43 (17.5%)	259	41 (15.8%)	0.705[e]
				bd[d]	240	36 (15.0%)			0.895[e]
		CIBIC-plus: no change – 26wk	D	od[b]	246	90 (36.6%)	259	94 (36.3%)	0.981[e]
				bd[d]	240	93 (38.8%)			0.636[e]
		CIBIC-plus: minimally worse – 26wk	D	od[b]	246	69 (28.0%)	259	70 (27.0%)	0.875[e]
				bd[d]	240	67 (27.9%)			0.903[e]
		CIBIC-plus: moderately worse – 26wk	D	od[b]	246	23 (9.3%)	259	36 (13.9%)	0.146[e]
				bd[d]	240	25 (10.4%)			0.294[e]
		CIBIC-plus: markedly worse – 26wk	D	od[b]	246	4 (1.6%)	259	6 (2.3%)	0.812[e]
				bd[d]	240	2 (0.8%)			0.336[e]
Rockwood et al. (2006)[125]	LOCF analysis	CIBIC-plus score – 8wk	A		61	3.64 (SD 0.797)	65	4.17 (SD 0.905)	-
		CIBIC-plus score – 16wk	A		61	3.67 (SD 0.996)	65	4.12 (SD 0.987)	0.03[f]

[a] A=absolute value; D=dichotomous

[b] galantamine prolonged release formulation once a day

[c] Cochrane-Mantel-Haenszel statistic using modified ridit scores, derived from rank score (the Van Elteren test) and controlling for country effect (United States v. ex-United States)

[d] galantamine standard formulation twice a day

[e] chi-square test (Yates's correction) (calculated by reviewer)

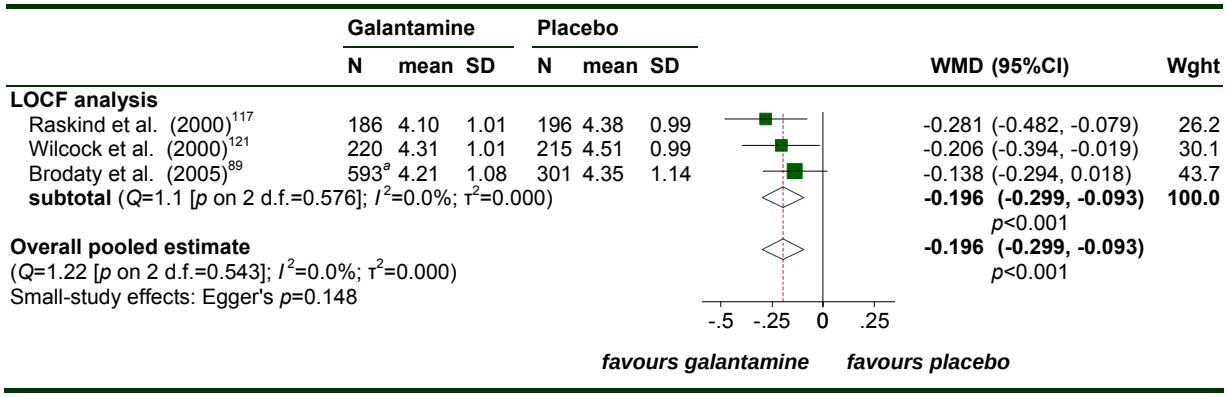
[f] test not stated

Synthesis with existing evidence-base

Clinician Interview-Based Impression of Change

When the new studies' data were pooled with the existing evidence base the overall pooled estimates of the CIBIC-plus at 26 weeks showed a benefit from galantamine compared with placebo, WMD=-0.20 (95%CI -0.30, -0.09), p<0.001 (*Figure 31*).

FIGURE 31 *Random-effects meta-analysis: CIBIC-plus at 26wk – galantamine (maximum dose ≤24mg/d) v. placebo*



[a] once-daily prolonged release formulation and twice-daily standard formulation pooled

4.6.2.2.4.1. Pooled multiple outcome measures

Because CIBIC-plus is the only outcome measure used to assess global effect in placebo-controlled trials of galantamine, there was no reason to pool outcomes on a standardised level for this domain.

4.6.2.2.5. Quality of life

None of the included studies provided any randomised evidence on QoL with galantamine compared with placebo, and no such data were identified in the 2004 review.

4.6.2.2.6. Safety

Overall there was a high percentage of any AE in both studies in treatment and control groups (any AE, Brodaty, once/day treatment =79%, placebo=70% and Rockwood, treatment=84%, placebo=62%). The main AEs were gastrointestinal. A summary of all the AEs reported can be found in *Table 16*.

TABLE 16 *AEs in included studies – galantamine* v. *placebo*

Adverse events	Galantamine od			Galantamine bd						Placebo			
	Brodaty et al. (2005)[89]			Brodaty et al. (2005)[89]			Rockwood et al. (2006)[125]			Brodaty et al. (2005)[89]		Rockwood et al. (2006)[125]	
	N	n (%)	*p*[a]	N	n (%)	*p*[a]	N	n (%)	*p*[a]	N	n (%)	N	n (%)
Any AE	319	253 (79.3%)	0.009	326	235 (72.1%)	0.619	64	54 (84.4%)	0.008	320	224 (70.0%)	66	41 (62.1%)
Any gastrointestinal	319	111 (34.8%)	0.009	326	114 (35.0%)	0.007				320	80 (25.0%)		
Any psychiatric	319	73 (22.9%)	0.551	326	58 (17.8%)	0.415				320	66 (20.6%)		
Any general	319	76 (23.8%)	0.141	326	62 (19.0%)	0.989				320	60 (18.8%)		
Any central/peripheral nervous system	319	77 (24.1%)	0.017	326	69 (21.2%)	0.134				320	52 (16.3%)		
Any respiratory	319	45 (14.1%)	0.896	326	41 (12.6%)	0.835				320	43 (13.4%)		
Any metabolic/nutritional	319	42 (13.2%)	0.536	326	43 (13.2%)	0.527				320	36 (11.3%)		
Any urinary	319	40 (12.5%)	0.892	326	39 (12.0%)	0.931				320	38 (11.9%)		
Any secondary term	319	28 (8.8%)	0.201	326	30 (9.2%)	0.271				320	39 (12.2%)		
Anorexia	319	19 (6.0%)	0.048	326	22 (6.7%)	0.017	64	7 (10.9%)	0.061	320	8 (2.5%)	66	1 (1.5%)
Nausea	319	54 (16.9%)	0.000	326	45 (13.8%)	0.000	64	15 (23.4%)	0.011	320	16 (5.0%)	66	4 (6.1%)
Diarrhoea	319	15 (4.7%)	0.314	326	22 (6.7%)	0.926				320	22 (6.9%)		
Vomiting	319	21 (6.6%)	0.012	326	28 (8.6%)	0.001	64	11 (17.2%)	0.017	320	7 (2.2%)	66	2 (3.0%)
Agitation	319	22 (6.9%)	0.992	326	20 (6.1%)	0.951				320	21 (6.6%)		
Depression	319	18 (5.6%)	0.070	326	16 (4.9%)	0.159				320	8 (2.5%)		
Injury	319	24 (7.5%)	0.419	326	12 (3.7%)	0.324				320	18 (5.6%)		
Dizziness	319	33 (10.3%)	0.006	326	24 (7.4%)	0.148				320	14 (4.4%)		
Headache	319	29 (9.1%)	0.127	326	18 (5.5%)	0.909				320	18 (5.6%)		
Upper respiratory tract infection	319	15 (4.7%)	0.993	326	12 (3.7%)	0.529	64	8 (12.5%)	0.090	320	16 (5.0%)	66	2 (3.0%)
Weight decrease	319	14 (4.4%)	0.031	326	17 (5.2%)	0.009				320	4 (1.3%)		
Urinary tract infection	319	22 (6.9%)	0.661	326	22 (6.7%)	0.605				320	26 (8.1%)		
Fall	319	20 (6.3%)	0.992	326	20 (6.1%)	0.952				320	19 (5.9%)		

[a] all *p*-values represent galantamine *v.* placebo, as assessed by chi-squared test (Yates's correction), calculated by reviewer

4.6.2.3. Summary: galantamine v. placebo

We found an additional three RCTs to add to the five reported in the 2004 review.

Overall cognitive results from the new studies using ADAS-cog showed improvements for those taking galantamine. When these studies were pooled with the existing evidence the benefit remained and increased with time, with greater benefit seen at 21–26 weeks WMD=-2.96 (95%CI -3.41, -2.51), p<0.001) than 12–16 weeks WMD=-2.39 (95%CI -2.80, -1.97), p<0.001).

All the new studies reported functional outcomes. Those measured by the DAD and ADCS-ADL scales generally showed significant improvement, those measured by the Goal Attainment Scale (GAS) were rather more ambiguous. Pooled results of the ADCS-ADL and the DAD at 21–26 weeks continued to show benefit from galantamine compared to placebo (WMD=2.23 (95%CI 1.33, 3.14), p<0.001 and WMD=3.76 (95%CI 1.66, 5.86), p<0.001) respectively. When data from both these outcome measures were pooled, results still favoured galantamine.

Behavioural outcomes from one new study, measured by the NPI, failed to show a benefit from galantamine. This lack of benefit was also seen from the pooled results at 13 weeks from follow-up. However, when the new data were pooled with those of the previous assessment a significant difference favouring galantamine was found at 21–26 weeks (WMD=-1.46 (95%CI -2.59, -0.34), p=0.012).

Two of the new studies measured global outcomes; one found that it produced a significant benefit on the CIBIC-plus. When these data were pooled with the data from the previous review, significant benefit was found on the CIBIC-plus at 26 weeks follow-up WMD =-0.20 (95%CI -0.30, -0.09), p<0.001 with doses of ≤24mg/d.

No QoL data were reported in either the new or the old studies for this comparison. The main adverse events found were gastrointestinal.

The evidence from the new studies confirmed the benefits from galantamine for cognitive outcomes and although there was some ambiguity with functional measures the overall pooled estimates were favourable. While none of the new studies showed significant benefit for behavioural outcomes, when these data were pooled with existing evidence, gains were

shown at later follow-up times. Again, although the new trial data varied in their significance, pooling suggested that there were gains to be made on global outcomes. However, it should be noted that in all these studies estimates for missing data were calculated using LOCF or OC methods, and may therefore have inflated the benefits from galantamine.

4.6.2.4. Graphical summary of galantamine v. placebo

In contrast to the donepezil research, one large study has been conducted since 2004 (comparing two different methods of delivery v. placebo), and the quality of the studies has improved overall (see *Figure 32* and *Figure 33*). As in the previous review, the larger studies are more likely to show a benefit from galantamine on cognitive outcomes. The evidence for an effect on functional outcomes continues to be mixed; previously there appeared to be a relationship to the size of the dose, in the new study by Brodaty and colleagues[89] effectiveness at 24 mg/day was linked to mode of delivery, with the prolonged release capsule (PRC) showing a significant gain over placebo, which the regular capsule did not. Similar to the 2004 review the results on global out comes were mixed; thus it remains unclear what effect galantamine has on these outcomes.

FIGURE 32 *Summary of studies included in the 2004 review – galantamine v. placebo*

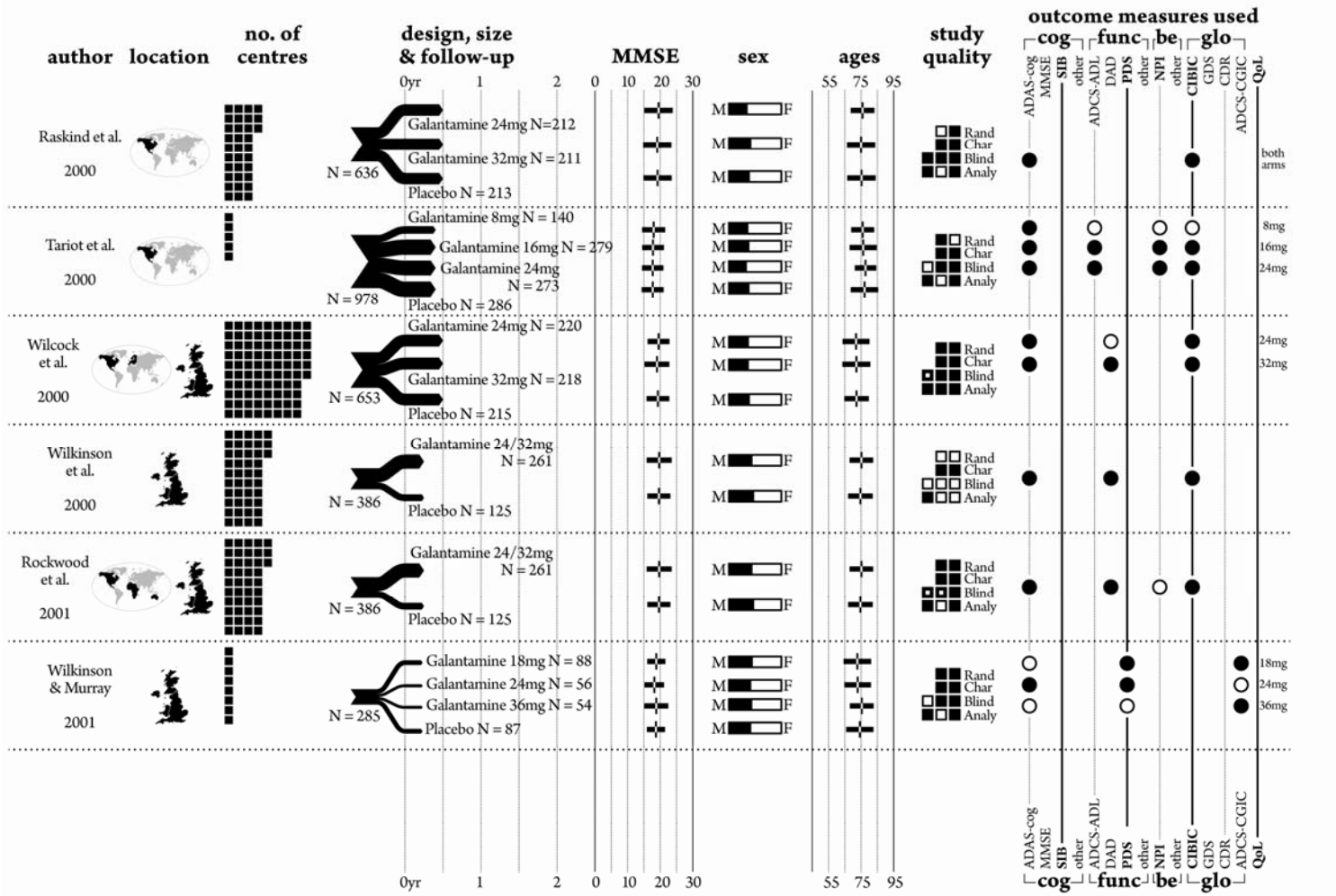
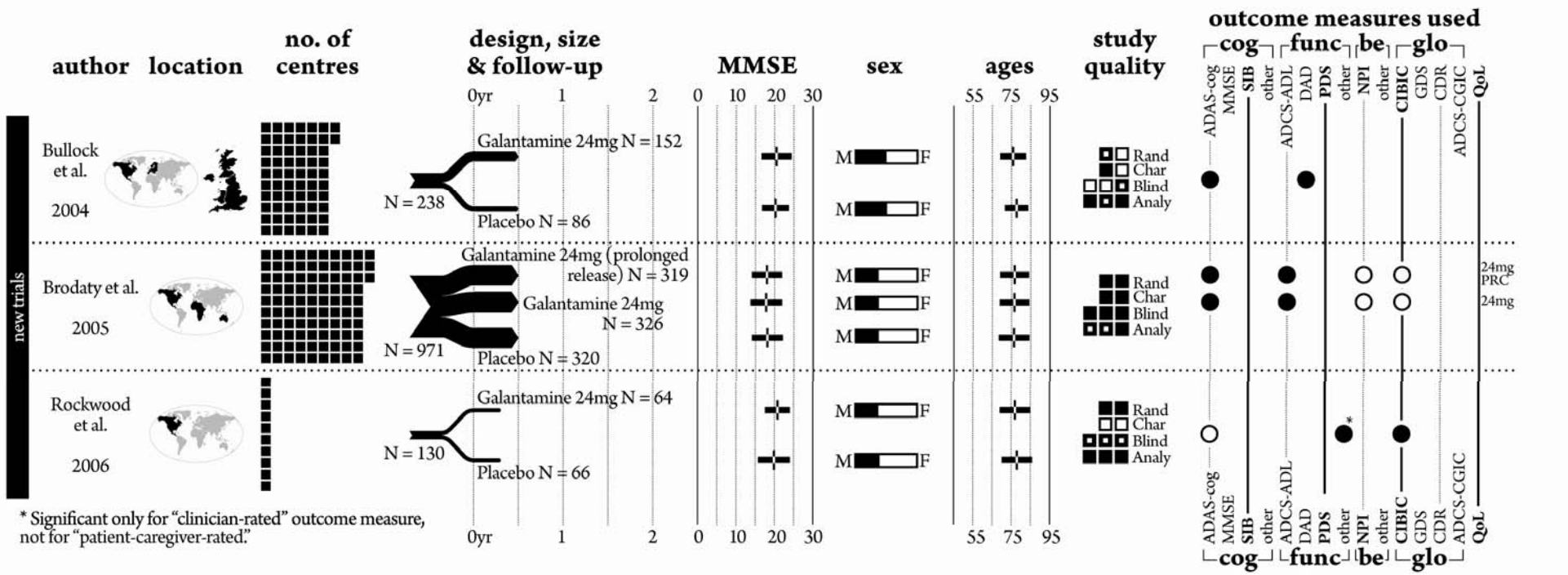


FIGURE 33 *Summary of new studies included in the 2010 review – galantamine v. placebo*



4.6.3. Rivastigmine *v.* placebo

4.6.3.1. Identified evidence

The 2004 review included four RCTs, those reported by Agid and colleagues. (1998),[128] Corey-Bloom and colleagues. (1998),[129] Forette and colleagues. (1999),[130] and Rosler and colleagues. (1999).{Rosler, 1999 2016 /id}

Our searches identified three additional relevant trials,[132-134] details of which are tabulated in *Table 17*. *Table 18* contains information about studies' interventions, comparators and baseline characteristics and *Table 19* gives key markers of internal validity for the included studies, which were moderately good to poor quality.

TABLE 17 *Design of included studies – rivastigmine* v. *placebo*

Study details	Inclusion criteria	Exclusion criteria	Methodological notes	Other
Feldman & Lane (2007)[132] **Design:** Parallel double-blind RCT **Country:** Australia, Canada, Ireland, Italy, South Africa, UK **No. of centres:** 37 **No. randomised:** 678 **Maximum follow-up:** 26 **MMSE range included:** 10–26	AD (DSM-IV criteria) and probable AD (NINCDS-ADRDA) Responsible caregiver	Severe and unstable cardiac disease Severe and obstructuive pulmonary disease Other life-threatening conditions Use of anticholoinergic drugs, health food supplements containing ACh precursors, putative memory enhancers, or insulin Use of psychotropic drugs, with the exception of chloral hydrate, short acting benzodiazepines and haloperidol (<=3d in succession and not <72h before any efficacy assessment)	**Sample attrition / dropout:** 553 of 678 completed study. 125 withdrew after allocation: adverse events (n=83); ECG abnormalities (n=4); laboratory abnormalities (n=1); withdrawn consent (n=14); protocol violation (n=8); treatment failure (n=2); failure to attend (n=7); other reasons (n=6). Differences between groups was only on adverse events (rivastigmine TID 11%; rivastigmine BID 17%; placebo 9%) **Randomisation and allocation:** Randomisation procedure not described. Rivastigmine and placebo tablets were identical and the number taken was the same at each dose in all groups. **Power calculation:** The study sample size was determined on the basis of an estimated 3.0 point difference between rivastigmine administered BID and placebo on the ADAS-cog, an estimated 0.4 point difference between BID and placebo on the CIBIC-Plus and an	**Therapy common to all participants:** None **Study Funding:** Commissioned by Novartis Pharma AG (Switzerland) **Other conflicts:** HF has received honoraria for consulting, advisory boards and for participation in CME programs sponsored by Novartis. He has also received grant-in-aid funding for research from Novartis. RL is an employee of Novartis. The study was commissioned by Novartis Pharma AG in Switzerland.

Study details	Inclusion criteria	Exclusion criteria	Methodological notes	Other
			increased proportion of responders with CIBIC-Plus ratings of .4 of 20% within the BID rivastigmine group (35% rivastigmine vs 15% placebo). Sample sizes of 192 per group were required. For practical reasons the sample size was chosen as 200 (intention to treat (ITT) population). An individual power of 90% guaranteed protection of the global power in view of the requirement that both ADAS-cog and CIBIC-Plus analyses should be significant at the 0.0499 level.	
Mowla et al. (2007)[133] **Design:** Parallel double-blind RCT **Country:** Not reported. Lead author based in Iran **No. of centres:** Not reported **No. randomised:** 122 **Maximum follow-up:** 12 **MMSE range included:** 10–24	AD (DSM-IV criteria) Brief Cognitive Rating Score mean 3-5 Hachinski Iscahemic Score <4 Adequate level of premorbid intelligence (IG >80, global assessment)	Dementia of other aetiology Severe organic disease (tumours, severe infectious disease, brain trauma, epilepsy, cerebrovascular malformations, alcohol or drug abuse) Other psychiatric disorders (Hamilton Depression Scale, 17-item version, total score <10)	**Sample attrition / dropout:** 98 of 122 completed study. Drop-outs: Rivastigmine arm n=7; Fluoxetine plus rivastigmine n=9; placebo n=8. Major cause of withdrawal in fluoxetine plus rivastigmine arm was adverse events, in placebo arm it was loss of efficacy. **Randomisation and allocation:** Computer-generated (on-site) randomisation – whether researchers were able to view randomisation sequence prior to allocation is not reported. Same number of pills for all trial arms, but appearance of these pills not reported (simply described as 'similar') **Power calculation:** Not reported	**Therapy common to all participants:** Single-blind placebo 6-week run-in period to exclude placebo responders **Study Funding:** Shiraz University of Medical Sciences **Other conflicts:** Not reported **Notes:** 12-week mean MMSE/WMS/ADL/HAM scores in the fluoxetine plus rivastigmine arm were much lower than in the other arms – potential error?
Winblad et al. (2007)[134] **Design:** Parallel double-blind RCT **Country:** Chile, Czech Republic, Denmark, Finland, Germany, Guatemala, Israel, Italy, Korea, Mexico, Norway, Peru, Poland, Portugal, Russia, Slovak	AD (DSM-IV criteria) and probable AD (NINCDS/ADRDA criteria) (brain scan (MRI or CT) used for establishing these criteria must have been done within one year prior to randomization) Age 50-85yr Living with someone in the community or, if living alone, in daily contact with a	Advanced, severe, progressive, or unstable disease of any type that could interfere with study assessments or put the patient at special risk Any condition other than AD that could explain the dementia Use of any investigational drugs, new	**Sample attrition / dropout:** 970 of 1195 patients completed study. Reasons for drop-out: adverse events, withdrawn consent, lost to follow-up, death, unsatisfactory therapeutic effect. No difference between groups. **Randomisation and allocation:** Automated random assignment of treatment using an interactive voice-response system. Blocking was	**Therapy common to all participants:** None reported **Study Funding:** Novartis Pharma AG, Basel, Switzerland **Other conflicts:** 3 co-authors (SZ, JN, RL) are employees of Novartis. Remaining authors were investigators (BW, NA, GG, MO, CS) and/or Study Publication Committee members (BW, JC, NA, GG, MO, SZ, JN, RL). BW, JC, NA, GG, MO and CS have provided consultation services to many pharmaceutical companies that develop

Study details	Inclusion criteria	Exclusion criteria	Methodological notes	Other
Republic, Sweden, Taiwan, USA, Uruguay, Venezuela **No. of centres:** 100 **No. randomised:** 1195 **Maximum follow-up:** 24 **MMSE range included:** 10–20	responsible caregiver	psychotropic or dopaminergic agents, cholinesterase inhibitors or anti-cholinergic agents during the 4 weeks prior to randomization	done on a study centre basis. All personnel directly involved in the conduct of the study remained unaware of the active treatment groups until all data had been retrieved and finalized for analysis. Appearance of tablets, patches and placebo not reported. **Power calculation:** In previous placebo-controlled trials of the rivastigmine capsule in AD patients, a treatment difference to placebo in the ADAS-Cog change from baseline of approximately 2.5 points was observed in the Intent-to-Treat (ITT) analysis. In the current trial, a non-inferiority margin was pre-defined as 1.25 points on the ADAS-Cog to preserve 50% of this effect, which was considered the smallest value that could represent a clinically meaningful difference. To determine the power of this study, the assumptions on delta (difference in means) and standard deviation (SD) for the change in ADAS-Cog and ADCS-CGIC from baseline were based on 24 week data from the rivastigmine capsule studies that used the ADAS-Cog and CIBICplus. The ADCS-CGIC scale is comparable to the CIBIC-plus, which was used in previous rivastigmine capsule studies. To ensure that the study had adequate power, 1,040 evaluable patients were needed. In order to reach an overall power of 80% for all of the first three hypotheses (which is defined as the product of the individual powers), the sample size was 260 patients per treatment group.	dementia drugs, including Novartis. A writing committee prepared an initial draft of the manuscript, based on a report provided by Novartis, and all authors contributed to its finalization through interactive review. Data were collected by investigators and co-investigators, entered into a central database using electronic data capture software, and analyzed by Novartis Pharma AG, which vouches for the data and the analysis.

TABLE 18 *Interventions, comparators, and baseline characteristics of participants in included studies – rivastigmine* v. *placebo*

Study	Arm	Dose (mg/d)	Dosage details	N	Age	Sex (n male)	Race (n white)	Weight (kg)	Education (yrs)	Duration of dementia (mo)	ADAS-cog	MMSE
Feldman & Lane (2007)[132]	Rivastigmine td	2–12	Dose administered three times a day. Titrated from an initial dose of 2mg/d for the first week up to a maximum of 12mg in 1mg/d steps at weekly intervals. Patients unable to tolerate 2mg/d by day 10 were withdrawn from the study. Tolerability could be optimised by maintaining a dose level for periods of up to 2wk.	227	71.4 (SD 7.90)	91 (40.1%)[a]		65.9 (SD 12.9)		38.4 (SD 25.5)	28.1 (SD 12.5)	18.3 (SD 4.50)
	Rivastigmine bd	2–12	Dose administered two times a day (plus one placebo tablet). Titrated from an initial dose of 2mg/d for the first week up to a maximum of 12mg in 1mg/d steps at weekly intervals. Patients unable to tolerate 2mg/d by day 10 were withdrawn from the study. Tolerability could be optimised by maintaining a dose level for periods of up to 2wk.	229	71.0 (SD 8.20)	98 (42.8%)[a]		66.7 (SD 12.2)		40.6 (SD 31.2)	27.7 (SD 12.3)[b]	18.8 (SD 4.60)
	Placebo	-	-	222	71.7 (SD 8.70)	89 (40.1%)[a]		65.9 (SD 12.3)		39.7 (SD 28.2)	28.5 (SD 12.3)[c]	18.7 (SD 4.60)
Mowla et al. (2007)[133]	Rivastigmine	3–12	Titrated from initial dose of 1.5mg twice a day, doubled every 2wk until maximum dose of 6mg twice a day reached (or dose which patient could tolerate) no details of placebo fluoxetine administration	41	69.2[d]	65 (53.3%)[e,f]						16.3 (SD 4.10)
	Rivastigmine+ Fluoxetine	3–12	Titrated from initial dose of 1.5mg twice a day, doubled every 2wk until maximum dose of 6mg twice a day reached (or dose which patient could tolerate) Fluoxetine 20mg/d	41	69.2[d]							15.6 (SD 0.730)
	Placebo	-	-	40	69.2[d]							16.5 (SD 3.60)
Winblad	~~Rivastigmine~~	~~4.75–~~	~~Titrated from initial 5cm^2 dose up to~~	~~291~~	~~73.6~~	~~93~~	~~220~~		~~9.90~~	~~13.2~~	~~27.0~~	~~16.6~~

Study	Arm	Dose (mg/d)	Dosage details	N	Age	Sex (n male)	Race (n white)	Weight (kg)	Education (yrs)	Duration of dementia (mo)	ADAS-cog	MMSE
	patch (10cm^2)	9.5	10cm^2 patch in 5cm^2 step at 4wk interval, followed by an 8wk maintenance phase Dose adjustments permitted to address perceived safety or tolerability issues. Placebo capsules administered according to regimen for active capsule arm.		(SD 7.90)	(32.0%)	(75.6%)		(SD 4.30)	(SD 16.8)	(SD 10.3)[g]	(SD 3.10)
et al. (2007)[134]	Rivastigmine patch (20cm^2)	4.75–17.4	Titrated from initial 5cm^2 dose up to 20cm^2 patch in 5cm^2 steps at 4wk intervals, followed by an 8wk maintenance phase Dose adjustments permitted to address perceived safety or tolerability issues. Placebo capsules administered according to regimen for active capsule arm.	302	74.2 (SD 7.70)	103 (34.1%)[f]	227 (75.2%)		9.90 (SD 4.40)	13.2 (SD 16.8)	27.4 (SD 9.70)[h]	16.6 (SD 2.90)
	Rivastigmine capsules	3–12	Initial dosage of 3mg/dy titrated upwards in steps of 3mg/dy up to a maximum of 12mg/dy Dose adjustments permitted to address perceived safety or tolerability issues. Placebo patch.	294	72.8 (SD 8.20)	101 (34.4%)	219 (74.5%)		9.90 (SD 4.40)	13.2 (SD 16.8)	27.9 (SD 9.40)[i]	16.4 (SD 3.10)
	Placebo	-	placebo capsules + placebo patch	302	73.9 (SD 7.30)	101 (33.4%)	227 (75.2%)		9.90 (SD 4.30)	13.2 (SD 16.8)	28.6 (SD 9.90)[j]	16.4 (SD 3.00)

[a] approximated to nearest integer (percentages only presented in text)

[b] *n*=227

[c] *n*=220

[d] average value for participants across all arms

[e] whole trial population (*n*=122)

[f] approximated to nearest integer (percentages only presented in text); poor rounding suggests true denominator may be less than full sample size

[g] *n*=248 (LOCF population)

[h] *n*=262 (LOCF population)

[i] *n*=253 (LOCF population)

[j] *n*=281 (LOCF population)

TABLE 19 *Markers of internal validity of included studies – rivastigmine* v. *placebo*

	Was the assignment to the treatment groups really random?	Was the treatment allocation concealed?	Were the groups similar at baseline in terms of prognostic factors?	Were the eligibility criteria specified?	Were outcome assessors blinded to the treatment allocation?	Was the care provider blinded?	Was the patient blinded?	Were the point estimates and measure of variability presented for the primary outcome measure?	Did the analyses include an intention-to-treat analysis?	Were withdrawals and dropouts completely described?
Feldman & Lane (2007)[132]	UNKNOWN	UNKNOWN	REPORTED – YES	UNKNOWN	UNKNOWN	ADEQUATE	ADEQUATE	ADEQUATE	ADEQUATE	ADEQUATE
Mowla *et al.* (2007)[133]	PARTIAL	ADEQUATE	REPORTED – YES	UNKNOWN	PARTIAL	PARTIAL	PARTIAL	ADEQUATE	INADEQUATE	ADEQUATE
Winblad *et al.* (2007)[134]	ADEQUATE	ADEQUATE	REPORTED – YES	ADEQUATE	ADEQUATE	PARTIAL	PARTIAL	ADEQUATE	ADEQUATE	ADEQUATE

4.6.3.2. Evidence of clinical effectiveness

4.6.3.2.1. Cognition

The 2004 assessment report by Loveman and colleagues summarised the evidence they found on cognitive outcomes as:

> *"Statistically significant differences between the 6-12mg/day treatment groups and placebo were reported by two of three published trials which reported ADAS-cog and MMSE. No statistically significant effects were seen in the low dose treatment groups in these studies. However, sample sizes were very low (<30 participants in each group) and this study presented no information on power calculations.*
>
> *The unpublished studies both found statistically better mean changes from ADAS-cog baseline scores in participants taking rivastigmine compared with placebo groups, a statistically significantly higher percentage of participants receiving TiD (x3 daily). Rivastigmine showed an improvement of at least four points compared with the placebo group, but there was no statistically significant difference between the BiD (x2 daily) group and placebo. One of the studies also reported on ADAS-cogA, and found statistically significant differences in both mean change from baseline and percentage of improvers for both BiD. and TiD. treatment groups compared with placebo.*
>
> *Both unpublished studies reported MMSE as an outcome measure and found a statistically significant improvement in participants receiving rivastigmine compared with those receiving a placebo, with the exception of the 9mg/day rivastigmine group."*[2]

New data

In the three studies we found that had been published since 2004, comparing rivastigmine with placebo for mild-to-moderate AD, all reported benefits for the treatment group on cognitive outcome measures. These results appear to be dose dependent (as in the previous report), with doses ≥ 12mg/day showing a greater effect (Appendix 5). Although only one study measured missing outcomes with an ITT population.[132] *Table 20* shows the summary results for cognition in the included studies. It should be noted that in the study by Winblad and colleagues[134] only the 10cm patch is currently licensed in the UK.

TABLE 20 *Measures of cognition in included studies – rivastigmine v. placebo*

Study	Subgroup	Outcome	Type[a]	Arm	Rivastigmine		Placebo		*p*
					N1	Mean1	N2	Mean2	
Feldman & Lane (2007)[132]	ITT population	ADAS-cog – 12wk	MC	bd	227	-1.9 (SD 6.66)[b]	220	0.9 (SD 5.93)[b]	<0.001[c]
				tid	228	-0.8 (SD 6.04)[b]			<0.05[c]
		ADAS-cog – 18wk	MC	bd	227	-1.6 (SD 6.66)[b]	220	1.8 (SD 6.67)[b]	<0.001[c]
				tid	228	-0.1 (SD 6.79)[b]			<0.001[c]
		ADAS-cog – 26wk	MC	bd	227	-0.2 (SD 7.3)	220	2.8 (SD 7.2)	≤0.001[d]
				tid	228	1.2 (SD 7.2)			<0.05[d]
		ADAS-cog: any improvement – 12wk	D	bd	227	68 (30.0%)[b]	220	36 (16.4%)[b]	≤0.001[e]
				tid	228	52 (22.8%)[b]			<0.05[e]
		ADAS-cog: any improvement – 18wk	D	bd	227	75 (33.0%)[b]	220	28 (12.7%)[b]	≤0.001[e]
				tid	228	57 (25.0%)[b]			≤0.001[e]
		ADAS-cog: any improvement – 26wk	D	bd	227	52 (22.9%)[b]	220	28 (12.7%)[b]	NS[e]
				tid	228	41 (18.0%)[b]			<0.05[e]
		ADAS-cogA – 26wk	MC	bd	227	-0.1 (SD 7.9)	220	3.2 (SD 7.8)	≤0.001[d]
				tid	228	1.5 (SD 7.8)			<0.05[d]
		MMSE – 26wk	MC	bd	227	0.3 (SD 3.6)	220	-1.4 (SD 3.6)	≤0.001[c]
				tid	227	-0.6 (SD 3.6)			<0.05[c]
	LOCF analysis	ADAS-cog – 26wk	MC	bd	209	-0.7 (SD 6.9)	208	2.7 (SD 6.8)	≤0.001[d]
				tid	199	0.8 (SD 6.9)			<0.05[d]
		ADAS-cogA – 26wk	MC	bd	209	-0.6 (SD 7.5)	208	3.1 (SD 7.4)	≤0.001[d]
				tid	199	1 (SD 7.5)			<0.05[d]
		MMSE – 26wk	MC	bd	193	0.4 (SD 3.4)	198	-1.4 (SD 3.5)	≤0.001[c]
				tid	186	-0.4 (SD 3.5)			<0.05[c]
	OC population	ADAS-cog – 26wk	MC	bd	180	-0.9 (SD 6.8)	183	2.1 (SD 6.8)	≤0.001[d]
				tid	173	0.9 (SD 7)			NS[d]
Mowla et al. (2007)[133]	OC population[f]	MMSE – 12wk	A		34	17.4 (SD 3.7)	32	16 (SD 3.7)	0.129[g]
			MC		34	1.1 (SD 1.4)	32	-0.5 (SD 0.5)	<0.001[h]
		WMS-III – 12wk	A		34	8.7 (SD 2.2)	32	7.5 (SD 1.4)	0.011[g]
			MC		34	0.97 (SD 1.7)	32	-0.66 (SD 1.1)	<0.001[h]
		CGIC: item 2 (cognitive) – 12wk	A		34	3.1 (SD 0.96)	32	3.7 (SD 0.67)	0.005[g]
Winblad et al. (2007)[134]	LOCF analysis	ADAS-cog – 16wk	MC	10[i]	248	-0.825 (SD 6.3)[b]	281	0 (SD 6.71)[b]	0.09[j]
				20[k]	262	-1.39 (SD 6.47)[b]			<0.05[j]
				oral	253	-0.5 (SD 6.36)[b]			NS[j]
		ADAS-cog – 24wk	MC	10[i]	248	-0.6 (SD 6.4)	281	1 (SD 6.8)	0.005[j]
				20[k]	262	-1.6 (SD 6.5)			<0.001[j]
				oral	253	-0.6 (SD 6.2)			0.003[j]
		MMSE – 24wk	MC	10[i]	250	1.1 (SD 3.3)	281	0 (SD 3.5)	0.002[l]
				20[k]	262	0.9 (SD 3.4)			<0.001[l]
				oral	256	0.8 (SD 3.2)			0.002[l]
		Ten-point clock-drawing test – 24wk	MC	10[i]	251	0.1 (SD 3.1)	269	-0.1 (SD 3.2)	0.08[l]
				20[k]	245	0.3 (SD 3.4)			0.08[l]
				oral	246	0.2 (SD 2.9)			0.15[l]
		Trail-making test – 24wk	MC	10[i]	241	-12.3 (SD 55.1)	258	7.7 (SD 56.6)	<0.001[j]
				20[k]	238	-6.5 (SD 55.9)			0.005[j]
				oral	240	-9.8 (SD 66.1)			<0.001[j]

[a] MC=mean change; A=absolute value; D=dichotomous

[b] estimated from figure

[c] t-test using pooled error term from ANCOVA/ANOVA (SAS Type III analysis)

[d] Mantel–Haenszel test blocking for centre

[e] Mantel–Haenszel test

[f] publication does not explicitly state population in which outcomes were measured; description of withdrawals gives the impression that data may represent final OC population, which is what we have assumed

[g] student's t-test (two-tailed) (calculated by reviewer)

[h] post-hoc Tukey test

[i] 10cm^2 rivastigmine patch – equivalent to 9.5mg/d

[j] two-way ANCOVA (explanatory variables: treatment, country, and baseline scores)

[k] 20cm^2 rivastigmine patch – equivalent to 17.4mg/d

[l] Cochran–Mantel–Haenszel van Elteren test using modified ridit scores stratified by country

Synthesis with existing evidence-base

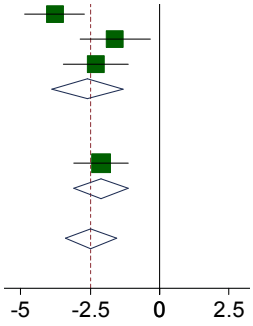
The data from the new trials were synthesised with those of the previous assessment using random-effects meta-analysis. The measures of ADAS-cog and MMSE were first considered separately and then combined in a pooled multiple outcome measure analysis at ≥12 mg/d. We also meta-analysed the data by ≤10 mg/d, 4 mg/d and combined doses these results can be found in Appendix 5.

ADAS-cog

The meta-analyses of ADAS-cog scores at 24–26 weeks showed a significant benefit from rivastigmine (≥12 mg/d) compared to placebo, WMD=-2.46 (95%CI-3.37, -1.56), p<0.001 (*Figure 33*).

FIGURE 34 *Random-effects meta-analysis: ADAS-cog at 24–26wk (mean change from baseline) – rivastigmine (≥12mg/d) v. placebo*

	Rivastigmine			Placebo				
	N	mean	SD	N	mean	SD	WMD (95%CI)	Wght
ITT population								
Corey-Bloom et al. (1998)[129]	231	0.31	5.97	234	4.09	6.01	-3.780 (-4.869, -2.691)	25.4
Rosler et al. (1999){Rosler, 1999 2016 /id}	242	-0.26	7.30	238	1.34	6.69	-1.600 (-2.853, -0.347)	22.8
Feldman & Lane (2007)[132]	455[a]	0.50	7.25	220	2.80	7.20	-2.298 (-3.460, -1.137)	24.2
subtotal (*Q*=7.19 [*p* on 2 d.f.=0.028]; I^2=72.2%; τ^2=0.917)							**-2.587 (-3.864, -1.311)**	**72.4**
							p<0.001	
LOCF analysis								
Winblad et al. (2007)[134]	515[b]	-1.11	6.35	281	1.00	6.80	-2.109 (-3.075, -1.143)	27.6
subtotal (*Q*=0.0 [*p* on 0 d.f.<0.001]; I^2=0.0%; τ^2=0.000)							**-2.109 (-3.075, -1.143)**	**27.6**
							p<0.001	
Overall pooled estimate							**-2.464 (-3.373, -1.555)**	
(*Q*=8.03 [*p* on 3 d.f.=0.045]; I^2=62.6%; τ^2=0.537)							*p*<0.001	
Inter-stratum heterogeneity: *p*=0.358								
Small-study effects: Egger's *p*=0.810								



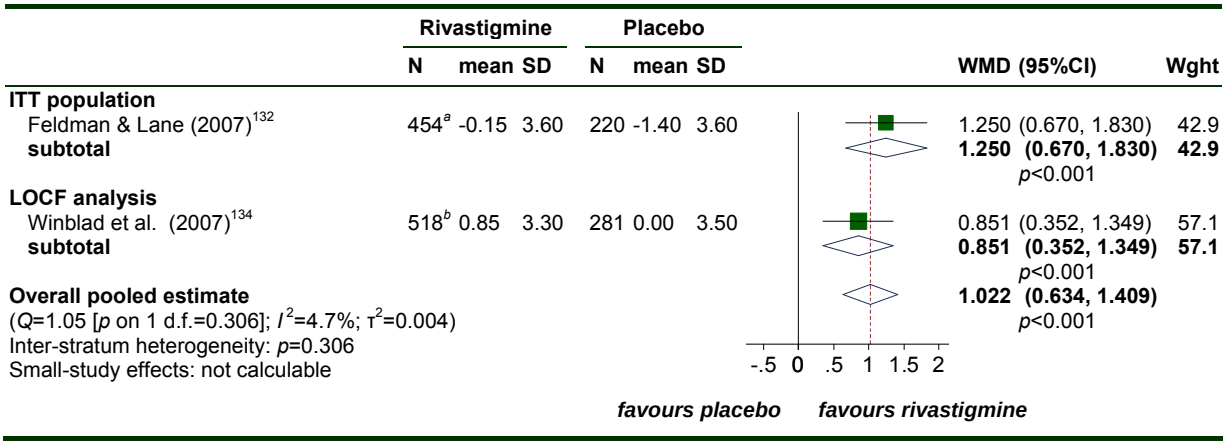
[a] bd and tid arms pooled

[b] 20cm^2 patch and 12mg/d capsules arms pooled

MMSE

At 24–26 weeks follow up the pooled estimate of effect showed a benefit from rivastigmine, WMD=1.02 (95%CI 0.63, 1.41), p<0.001 (*Figure 35*).

FIGURE 35 *Random-effects meta-analysis: MMSE at 24–26wk (mean change from baseline) – rivastigmine (≥12mg/d) v. placebo*



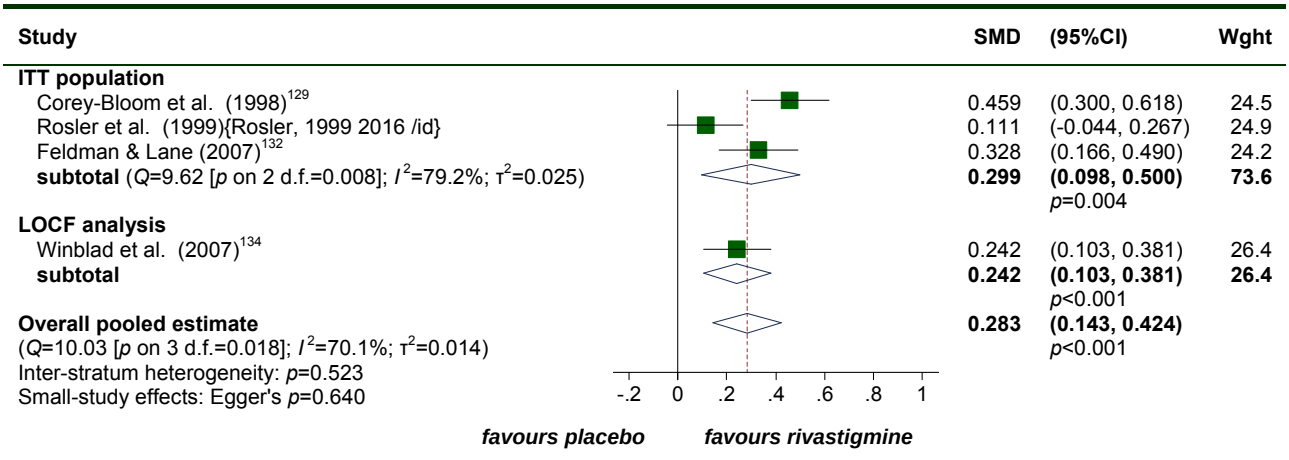
[a] bd and tid arms pooled

[b] 20cm^2 patch and 12mg/d capsules arms pooled

4.6.3.2.1.1. Pooled multiple outcome measures

When we pooled all the results for cognitive outcomes from the new and existing studies, we found that the overall pooled estimate showed a significant benefit from rivastigmine compared to placebo, SMD=0.28 (95%CI 0.14, 0.42),p<0.001 (*Figure 36*). The data set used in this meta-analysis can be found in Appendix 6.

FIGURE 36 *Random-effects meta-analysis: cognitive outcomes (SMD) at 24–26wk – rivastigmine (all dosages) v. placebo*



With only four datapoints in this evidence-base, it would not be informative to perform meta-regression.

4.6.3.2.2. Functional

In 2004 Loveman and colleagues reported that:

> *"Two published studies reported the PDS as a functional outcome measure. One of these found a statistically significant improvement in participants treated with 6-12mg/day rivastigmine compared with placebo, and the other reported that a statistically significantly higher percentage of these high dose participants than placebo participants showed an improvement of at least 10%."*[2]

New data

Two of the three new studies found since 2004 reported significant functional benefit from rivastigmine compared to placebo. These used the PDS and the ADCS-ADL as their outcome measures. A summary table of results can be found below in *Table 21*.

TABLE 21 *Measures of functional ability in included studies – rivastigmine v. placebo*

Study	Subgroup	Outcome	Type[a]	Arm	Rivastigmine		Placebo		p
					N	Mean	N	Mean	
Feldman & Lane (2007)[132]	ITT population	PDS – 26wk	MC	bd	225	-1.5 (SD 11.3)	221	-4.9 (SD 11.2)	≤0.001[b]
				tid	227	-2.6 (SD 11.1)	221	-4.9 (SD 11.2)	<0.05[b]
	LOCF analysis	PDS – 26wk	MC	bd	207	-1 (SD 11.4)	209	-4.7 (SD 11.3)	≤0.001[b]
				tid	195	-2.3 (SD 11.5)	209	-4.7 (SD 11.3)	<0.05[b]
Mowla et al. (2007)[133]	OC population	ADL – 12wk	A		34	25.3 (SD 6.6)	32	27.1 (SD 6.9)	0.283[c]
			MC		34	1.2 (SD 2.6)	32	-0.68 (SD 1.3)	0.58[d]
Winblad et al. (2007)[134]	LOCF analysis	ADCS-ADL – 16wk	MC	10[e]	247	-0.6 (SD 9.43)[f]	281	-1.6 (SD 7.96)[f]	NS[g]
				20[h]	263	0.4 (SD 9.73)[f]	281	-1.6 (SD 7.96)[f]	<0.05[g]
				oral	254	-0.4 (SD 7.97)[f]	281	-1.6 (SD 7.96)[f]	NS[g]
		ADCS-ADL – 24wk	MC	10[e]	247	-0.1 (SD 9.1)	281	-2.3 (SD 9.4)	0.01[g]
				20[h]	263	0 (SD 11.6)	281	-2.3 (SD 9.4)	0.02[g]
				oral	254	-0.5 (SD 9.5)	281	-2.3 (SD 9.4)	0.04[g]

[a] MC=mean change; A=absolute value

[b] Mantel–Haenszel test blocking for centre

[c] student's t-test (two-tailed) (calculated by reviewer)

[d] post-hoc Tukey test (NB *t*-test $p<0.001$)

[e] 10cm^2 rivastigmine patch – equivalent to 9.5mg/d

[f] data extracted from figure

[g] two-way ANCOVA (explanatory variables: treatment, country, and baseline scores)

[h] 20cm^2 rivastigmine patch – equivalent to 17.4mg/d

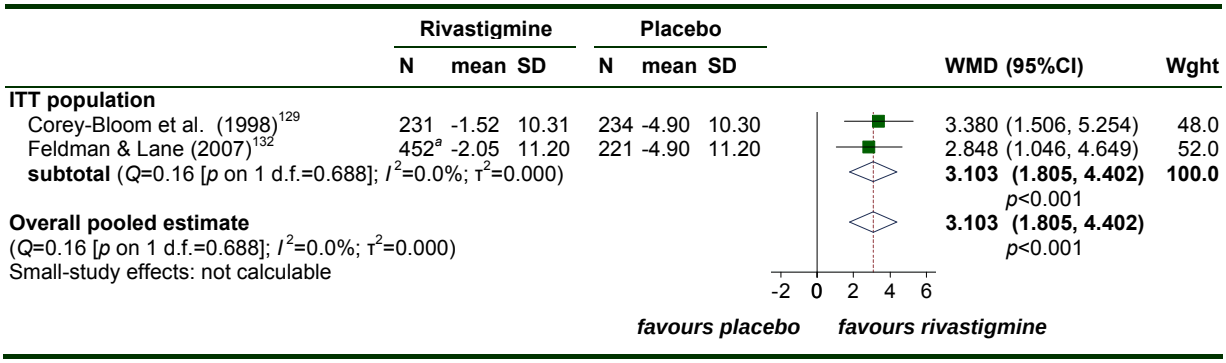
Synthesis with existing evidence-base

Data from the existing evidence was synthesised with the new data in a meta-analysis of the PDS.

Progressive Deterioration Scale

The overall pooled estimate at 24-26 weeks showed a significant benefit from rivastigmine, WMD=3.10 (95%CI 1.81, 4.40), p=0.001 (*Figure 37).*

FIGURE 37 *Random-effects meta-analysis: PDS at 24–26wk (mean change from baseline) – rivastigmine (12mg/d) v. placebo*

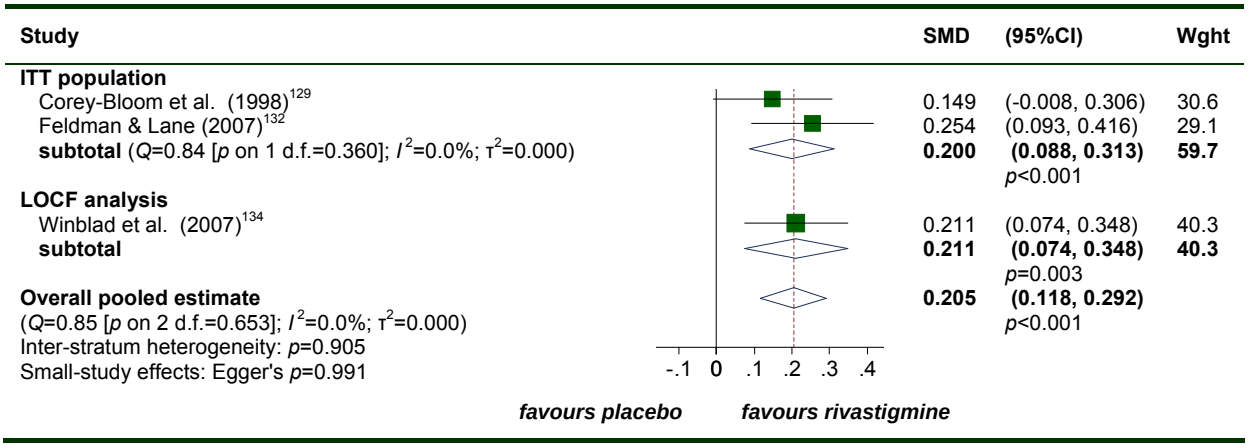


[a] bd and tid arms pooled

4.6.3.2.2.1. Pooled multiple outcome measures

Two new studies were found to add to this combined meta-analysis of functional outcomes at 24-26 weeks. Again, the overall pooled estimate showed a benefit from rivastigmine compared to placebo, SMD=0.21 (95%CI 0.12, 0.29), p<0.001 (*Figure 38*). The data set used in this meta-analysis can be found in Appendix 6.

FIGURE 38 *Random-effects meta-analysis: functional outcomes (SMD) at 24–26wk – rivastigmine (all dosages) v. placebo*



With only three datapoints in this evidence-base, it would not be informative to perform meta-regression.

4.6.3.2.3. Behavioural and mood

The 2004 systematic review summarised the behavioural results as:

> *"On measures of behaviour and mood no statistically significant benefit was demonstrated in the rivastigmine treated groups compared to the placebo groups."*[2]

New data

Two new studies were found that measured behavioural outcomes. One small study by Mowlal and colleagues found a significant benefit from rivastigmine[133] the other, much larger study, did not.[134] *Table 22* below shows the summary outcome data.

TABLE 22 *Measures of behavioural effect and mood in included studies – rivastigmine* v. *placebo*

Study	Subgroup	Outcome	Type[a]	Arm	Rivastigmine		Placebo		p
					N	Mean	N	Mean	
Mowla et al. (2007)[133]	OC population	Hamilton DS – 12wk	A		34	6.26 (SD 2.9)	32	8.33 (SD 1.12)	<0.001[b]
Winblad et al. (2007)[134]	LOCF analysis	NPI – 24wk	MC	10[c]	248	-1.7 (SD 11.5)	281	-1.7 (SD 13.8)	0.74[d]
		NPI – 24wk	MC	20[e]	263	-2.3 (SD 13.3)	281	-1.7 (SD 13.8)	0.69[d]
		NPI – 24wk	MC	oral	253	-2.2 (SD 11.9)	281	-1.7 (SD 13.8)	0.51[d]
		NPI – caregiver distress – 24wk	MC	10[c]	248	-1 (SD 5.5)	281	-1.1 (SD 6.3)	0.37[d]
		NPI – caregiver distress – 24wk	MC	20[e]	263	-1.1 (SD 6.4)	281	-1.1 (SD 6.3)	0.98[d]
		NPI – caregiver distress – 24wk	MC	oral	253	-1.1 (SD 6.6)	281	-1.1 (SD 6.3)	0.12[d]

[a] MC=mean change; A=absolute value

[b] student's *t*-test (calculated by reviewer)

[c] 10cm^2 rivastigmine patch – equivalent to 9.5mg/d

[d] two-way ANCOVA (explanatory variables: treatment, country, and baseline scores)

[e] 20cm^2 rivastigmine patch – equivalent to 17.4mg/d

Synthesis with existing evidence-base

The data identified by this review and the 2004 review are sparse and too heterogeneous to permit meaningful quantitative synthesis.

4.6.3.2.4. Global effect

The evidence from the 2004 assessment was summarised thus:

> *"Both of the published studies which included CIBIC-plus as a global outcome measure reported a statistically significant improvement in high dose participants (6-12mg/day) compared with placebo participants. One study also reported a statistically significantly greater proportion of 'responders' among participants treated with rivastigmine compared against placebo participants.*

> *Another study reported that a greater proportion of high dose rivastigmine participants than placebo participants had a 'successful' CIGIC assessment, i.e. scoring one or two on the scale. Two trials found a statistically significant improvement on the GDS measure in participants treated with 6-12mg/day of rivastigmine compared with placebo participants."*[2]

New data

The two new studies in this comparison that reported global outcomes had conflicting results. Feldman and Lane[132] found mostly significantly favourable results with the CIBIC-plus and the GDS, while Winblad and colleagues' results were mostly non-significant.[134] *(Table 23)*.

TABLE 23 *Measures of global effect in included studies – rivastigmine v. placebo*

Study	Subgroup	Outcome	Type[a]	Arm	Rivastigmine		Placebo		p
					N	Mean / n (%)	N	Mean / n (%)	
Feldman & Lane (2007)[132]	ITT population	CIBIC-plus score – 18wk	A	bd	220	3.9 (SD 1.04)[b]	213	4.5 (SD 1.02)[b]	≤0.001[c]
				tid	215	4.1 (SD 1.03)[b]			≤0.001[c]
		CIBIC-plus score – 26wk	A	bd	222	3.9 (SD 1.3)	216	4.5 (SD 1.3)	≤0.001[d]
				tid	222	4.1 (SD 1.3)			<0.05[d]
		CIBIC-plus: any improvement – 12wk	D	bd	220	66 (30.0%)[b]	213	34 (16.0%)[b]	≤0.001[e]
				tid	215	62 (28.8%)[b]			<0.05[e]
		CIBIC-plus: any improvement – 18wk	D	bd	220	68 (30.9%)[b]	213	40 (18.8%)[b]	≤0.001[e]
				tid	215	47 (21.9%)[b]			NS[e]
		CIBIC-plus: any improvement – 26wk	D	bd	220	68 (30.9%)[b]	213	40 (18.8%)[b]	<0.05[e]
				tid	215	49 (22.8%)[b]			NS[e]
		GDS – 26wk	MC	bd	227	0 (SD 0.7)	222	-0.3 (SD 0.7)	<0.05[c]
				tid	229	-0.2 (SD 0.7)			NS[c]
	LOCF analysis	CIBIC-plus score – 26wk	A	bd	206	3.9 (SD 1.2)	205	4.5 (SD 1.2)	≤0.001[d]
				tid	198	4.1 (SD 1.2)			<0.05[d]
		GDS – 26wk	MC	bd	195	0 (SD 0.7)	202	-0.3 (SD 0.7)	<0.05[c]
				tid	188	-0.1 (SD 0.7)			NS[c]
	OC population	CIBIC-plus score – 26wk	A	bd	177	3.9 (SD 1.2)	179	4.4 (SD 1.2)	≤0.001[d]
				tid	167	4.1 (SD 1.2)			<0.05[d]
Winblad et al. (2007)[134]	LOCF analysis	ADCS-CGIC: score – 16wk	A	10[f]	248	3.9 (SD 1.14)[g]	278	4.35 (SD 1.25)[g]	NS[h]
				20[i]	260	3.93 (SD 1.17)[g]			NS[h]
				oral	253	4.25 (SD 1.11)[g]			NS[h]
		ADCS-CGIC: score – 24wk	A	10[f]	248	3.9 (SD 1.2)	278	4.2 (SD 1.3)	0.01[h]
				20[i]	260	4 (SD 1.3)			0.054[h]
				oral	253	3.9 (SD 1.3)			0.009[h]
		ADCS-CGIC: markedly improved – 24wk	D	10[f]	248	5 (2.0%)	278	2 (0.7%)	0.361[j]
				20[i]	260	5 (1.9%)			0.395[j]
				oral	253	3 (1.2%)			0.916[j]
		ADCS-CGIC: moderately improved – 24wk	D	10[f]	248	29 (11.7%)	278	26 (9.4%)	0.463[j]
				20[i]	260	32 (12.3%)			0.334[j]
				oral	253	29 (11.5%)			0.513[j]
		ADCS-CGIC: minimally improved – 24wk	D	10[f]	248	43 (17.3%)	278	50 (18.0%)	0.937[j]
				20[i]	260	48 (18.5%)			0.975[j]
				oral	253	60 (23.7%)			0.129[j]
		ADCS-CGIC: unchanged – 24wk	D	10[f]	248	105 (42.3%)	278	91 (32.7%)	0.029[j]
				20[i]	260	94 (36.2%)			0.457[j]
				oral	253	96 (37.9%)			0.244[j]
		ADCS-CGIC: minimally worse – 24wk	D	10[f]	248	41 (16.5%)	278	65 (23.4%)	0.065[j]
				20[i]	260	50 (19.2%)			0.285[j]
				oral	253	30 (11.9%)			<0.001[j]
		ADCS-CGIC: moderately worse – 24wk	D	10[f]	248	22 (8.9%)	278	36 (12.9%)	0.177[j]
				20[i]	260	27 (10.4%)			0.429[j]
				oral	253	30 (11.9%)			0.803[j]
		ADCS-CGIC: markedly worse – 24wk	D	10[f]	248	3 (1.2%)	278	8 (2.9%)	0.303[j]
				20[i]	260	4 (1.5%)			0.448[j]
				oral	253	5 (2.0%)			0.696[j]

[a] MC=mean change; A=absolute value; D=dichotomous

[b] estimated from figure

[c] t-test using pooled error term from ANCOVA/ANOVA (SAS Type III analysis)

[d] t-test using pooled error term from ANOVA (SAS Type III)

[e] Mantel–Haenszel test

[f] 10cm^2 rivastigmine patch – equivalent to 9.5mg/d

[g] data extracted from figure

[h] Cochran–Mantel–Haenszel van Elteren test using modified ridit scores stratified by country

[i] 20cm^2 rivastigmine patch – equivalent to 17.4mg/d

[j] chi-square test (Yates's correction) (calculated by reviewer)

Synthesis with existing evidence-base

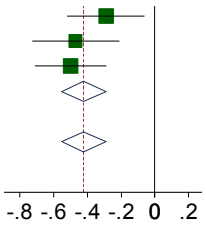
Data from the new studies were pooled with the existing evidence in random-effects meta-analyses using the CIBIC-plus at 26 weeks and the GDS at 26 weeks. The results can be seen in *Figure 39* and *Figure 40*.

Clinician Interview-Based Impression of Change

The meta-analysis showed a significant benefit from rivastigmine at 26 weeks, WMD = -0.42 (95%CI -0.55, -0.29), p<0.001.

FIGURE 39 *Random-effects meta-analysis: CIBIC-plus at 26wk – rivastigmine (12mg/d) v. placebo*

	Rivastigmine			Placebo				
	N	mean	SD	N	mean	SD	WMD (95%CI)	Wght
ITT population								
Corey-Bloom et al. (1998)[129]	231	4.20	1.24	234	4.49	1.25	-0.290 (-0.516, -0.064)	34.2
Rosler et al. (1999){Rosler, 1999 2016 /id}	219	3.91	1.51	230	4.38	1.24	-0.470 (-0.726, -0.214)	26.7
Feldman & Lane (2007)[132]	444[a]	4.00	1.30	216	4.50	1.30	-0.500 (-0.711, -0.289)	39.2
subtotal (*Q*=1.96 [*p* on 2 d.f.=0.374]; I^2=0.0%; τ^2=0.000)							**-0.420 (-0.553, -0.288)**	**100.0**
							p<0.001	
Overall pooled estimate							**-0.420 (-0.553, -0.288)**	
(*Q*=1.96 [*p* on 2 d.f.=0.374]; I^2=0.0%; τ^2=0.000)							*p*<0.001	
Small-study effects: Egger's *p*=0.974								



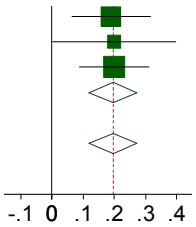
[a] bd and tid arms pooled

Global Deterioration Scale

This meta-analysis also showed a significant benefit from rivastigmine at 26 weeks, WMD 0.20 (95%CI 0.12, 0.27), p<0.001.

FIGURE 40 *Random-effects meta-analysis: GDS at 26wk (mean change from baseline) – rivastigmine (12mg/d) v. placebo*

	Rivastigmine			Placebo				
	N	mean	SD	N	mean	SD	WMD (95%CI)	Wght
ITT population								
Corey-Bloom et al. (1998)[129]	231	-0.13	0.70	234	-0.32	0.70	0.190 (0.063, 0.317)	37.1
Rosler et al. (1999){Rosler, 1999 2016 /id}	242	-0.06	1.11	238	-0.26	1.10	0.200 (0.002, 0.398)	15.3
Feldman & Lane (2007)[132]	456[a]	-0.10	0.70	222	-0.30	0.70	0.200 (0.087, 0.312)	47.6
subtotal (*Q*=0.01 [*p* on 2 d.f.=0.993]; I^2=0.0%; τ^2=0.000)							**0.196 (0.119, 0.274)**	**100.0**
							p<0.001	
Overall pooled estimate							**0.196 (0.119, 0.274)**	
(*Q*=0.01 [p on 2 d.f.=0.993]; I^2=0.0%; τ^2=0.000)							p<0.001	
Small-study effects: Egger's *p*=0.918								

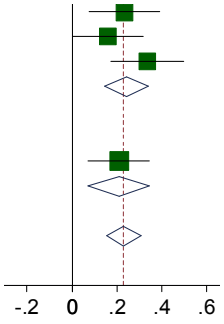


[a] bd and tid arms pooled

4.6.3.2.4.1. Pooled multiple outcome measures

We then pooled the results from both outcomes; the results from this can be seen in *Figure 41* and showed an overall pooled estimate of SMD=0.23 (95%CI 0.16, 0.31), p<0.001. The data set that was used in this meta-analysis can be found in Appendix 6.

FIGURE 41 *Random-effects meta-analysis: global outcomes (SMD) at 24–26wk – rivastigmine (all dosages) v. placebo*



Study	SMD	(95%CI)	Wght
ITT population			
Corey-Bloom et al. (1998)[129]	0.235	(0.078, 0.393)	23.5
Rosler et al. (1999){Rosler, 1999 2016 /id}	0.161	(0.003, 0.318)	23.6
Feldman & Lane (2007)[132]	0.334	(0.171, 0.496)	22.1
subtotal (*Q*=2.27 [*p* on 2 d.f.=0.322]; I^2=11.9%; τ^2=0.001)	**0.242**	**(0.144, 0.339)**	**69.2**
		p<0.001	
LOCF analysis			
Winblad et al. (2007)[134]	0.208	(0.071, 0.346)	30.8
subtotal	**0.208**	**(0.071, 0.346)**	**30.8**
		p=0.003	
Overall pooled estimate	**0.231**	**(0.155, 0.307)**	
(*Q*=2.42 [*p* on 3 d.f.=0.489]; I^2=0.0%; τ^2=0.000)		*p*<0.001	
Inter-stratum heterogeneity: *p*=0.695			
Small-study effects: Egger's *p*=0.575			

With only four datapoints in this evidence-base, it would not be informative to perform meta-regression.

4.6.3.2.5. Quality of life

None of the included studies provided any randomised evidence on QoL with rivastigmine compared with placebo, and no such data were identified in the 2004 review.

4.6.3.2.6. Safety

Overall there were a high percentage of any AEs, ranging from 51% to 91% in the treatment groups, and 46% to 76% in control groups. The main AEs were gastrointestinal, the lower dose (9.5 mg/day) transdermal patch produced fewer side effects than the capsule (12 mg/day). A summary of all the AEs reported can be found in *Table 24*.

TABLE 24 *AEs in included studies – rivastigmine* v. *placebo*

Adverse event	Rivastigmine td (≤12mg/d)			Rivastigmine bd (≤12mg/d)						Rivastigmine patch 10cm^2			Rivastigmine patch 20cm^2			placebo			
	Feldman & Lane (2007)[132]			Feldman & Lane (2007)[132]			Winblad et al. (2007)[134]			Winblad et al. (2007)[134]			Winblad et al. (2007)[134]			Feldman & Lane (2007)[132]		Winblad et al. (2007)[134]	
	N	n (%)	p[a]	N	n (%)	p	N	n (%)	p	N	n (%)	p	N	n (%)	p	N	n (%)	N	n (%)
Any AE	227	208 (91.6%)	<0.05[b]	228	208 (91.2%)	<0.05[b]	294	186 (63.3%)	≤0.001[c]	291	147 (50.5%)	NS[c]	303	200 (66.0%)	≤0.001[c]	222	169 (76.1%)	302	139 (46.0%)
Any serious AE	227	40 (17.6%)	NS[b]	228	40 (17.5%)	NS[b]										222	33 (14.9%)		
Anorexia	227	42 (18.5%)	<0.05[b]	228	47 (20.6%)	<0.05[b]										222	6 (2.7%)		
Nausea	227	109 (48.0%)	<0.05[b]	228	123 (53.9%)	<0.05[b]	294	68 (23.1%)	≤0.001[c]	291	21 (7.2%)	NS[c]	303	64 (21.1%)	≤0.001[c]	222	31 (14.0%)	302	15 (5.0%)
Diarrhoea	227	38 (16.7%)	<0.05[b]	228	40 (17.5%)	<0.05[b]	294	16 (5.4%)	NS[c]	291	18 (6.2%)	NS[c]	303	31 (10.2%)	≤0.001[c]	222	20 (9.0%)	302	10 (3.3%)
Vomiting	227	68 (30.0%)	<0.05[b]	228	88 (38.6%)	<0.05[b]	294	50 (17.0%)	≤0.001[c]	291	18 (6.2%)	NS[c]	303	57 (18.8%)	≤0.001[c]	222	14 (6.3%)	302	10 (3.3%)
Abdominal pain	227	26 (11.5%)	<0.05[b]	228	34 (14.9%)	<0.05[b]										222	12 (5.4%)		
Agitation	227	14 (6.2%)	<0.05[b]	228	21 (9.2%)	NS[b]										222	26 (11.7%)		
Anxiety	227	8 (3.5%)	NS[b]	228	13 (5.7%)	<0.05[b]										222	3 (1.4%)		
Dizziness	227	39 (17.2%)	<0.05[b]	228	42 (18.4%)	<0.05[b]	294	22 (7.5%)	≤0.01[c]	291	7 (2.4%)	NS[c]	303	21 (6.9%)	≤0.05[c]	222	16 (7.2%)	302	7 (2.3%)
Headache	227	36 (15.9%)	NS[b]	228	40 (17.5%)	<0.05[b]	294	18 (6.1%)	≤0.01[c]	291	10 (3.4%)	NS[c]	303	13 (4.3%)	NS[c]	222	23 (10.4%)	302	5 (1.7%)
Flatulence	227	15 (6.6%)	<0.05[b]	228	11 (4.8%)	NS[b]										222	4 (1.8%)		
Haemorrhoids	227	2 (0.9%)	NS[b]	228	0 (0.0%)	<0.05[b]										222	6 (2.7%)		
Weight loss							294	16 (5.4%)	≤0.01[c]	291	8 (2.7%)	NS[c]	303	23 (7.6%)	≤0.001[c]	302	4 (1.3%)		
Decreased appetite							294	12 (4.1%)	≤0.05[c]	291	2 (0.7%)	NS[c]	303	15 (5.0%)	≤0.01[c]	302	3 (1.0%)		
Asthenia							294	17 (5.8%)	≤0.001[c]	291	5 (1.7%)	NS[c]	303	9 (3.0%)	NS[c]	302	3 (1.0%)		

[a] all *p*-values represent rivastigmine *v.* placebo

[b] Fisher's exact test

[c] test not specified

4.6.3.3. Summary: rivastigmine v. placebo

Our update searches identified three new RCTs to add to the four included in the previous review.

All three studies showed benefits from rivastigmine on the ADAS-cog and MMSE, although these benefits were dependent on dose; with greater benefits seen at 12 mg/day than 6 mg/day. When these data were pooled with the existing evidence, significant differences favouring rivastigmine continued to be seen on the ADAS-cog at 24–26 weeks (≥12 mg/d), WMD=-2.46 (95%CI-3.37, -1.56) p<0.001. However, the benefits from rivastigmine were not apparent on MMSE scores until 24–26 weeks follow-up WMD=1.02 (95%CI 0.63, 1.41) p<0.001, this may be due to the MMSE's difficulties with detecting change. When the outcomes from both cognitive measures were combined they continued to show an advantage from taking rivastigmine on cognitive outcomes.

Two of the three new studies reporting functional outcomes showed significant gains for these measures. When these new data were synthesised with existing evidence using the PDS, significant gains were shown at 24–26 weeks, WMD=3.10 (95%CI 1.81, 4.40), p=0.001.

The data on behavioural outcomes from the new studies were unclear, with the smaller study showing a benefit from rivastigmine that the larger one did not. The existing evidence was too heterogeneous for meta-analysis, so the overall effectiveness of rivastigmine for behavioural outcomes is unknown.

The results for global outcomes were also mixed. Results from the CIBIC-plus were almost universally significant, whilst those measured by the ADCS-CGIC were almost universally not; those using the GDS showed no significant gain from rivastigmine. However, when these data were pooled with the existing evidence, the overall estimates favoured rivastigmine on the CIBIC-plus, WMD = -0.42 (95%CI -0.55, -0.29) p<0.001 and the GDS, WMD 0.20 (95%CI 0.12, 0.27) p<0.001. When results from both these outcome measures were combined, the result continued to show significant benefit from rivastigmine, these results are based on a robust ITT population.

When rivastigmine patches were compared to capsules, the results showed that the 9.5mg/day patch was similarly effective as the 12.5 mg/day capsule but with fewer side-effects.

None of the included studies in either the updated or original review reported QoL outcomes.

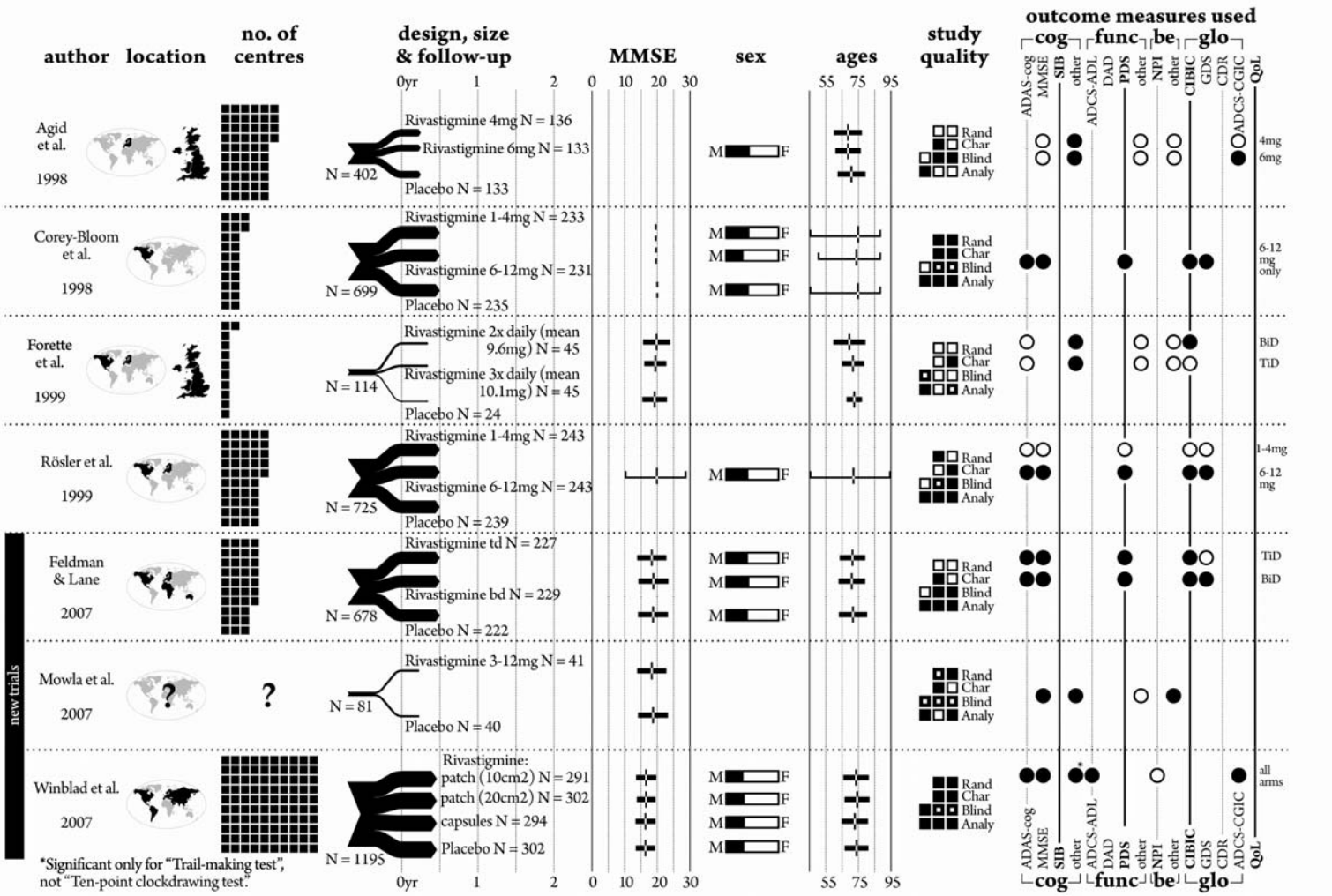
As in the other AChEIs, the main adverse events were gastrointestinal.

Overall, pooled estimates of cognitive benefits from rivastigmine were favourable, but were shown to be dose dependent as in the previous review. The results from functional and global outcomes also showed significant gains. However, results from individual trials of behavioural outcomes were mixed (pooling was not possible due to heterogeneity). The lower dose transdermal patch (9.5 mg/day) was shown to be as effective as the capsule (12 mg/day) but with fewer side-effects.

4.6.3.4. Graphical summary of rivastigmine v. placebo

The graphical summary in *Figure 42* shows that two large and one small study have been added to the evidence for rivastigmine since 2004. The graphic illustrates how the new studies have added to the precision of our knowledge of the effects of rivastigmine in AD. Previously the results for cognitive outcomes were ambiguous; however, the results from the new trials all show cognitive benefits. The smaller new study showed a gain on behavioural outcomes that had not been seen in the previous studies but the results for functional and global outcomes continued to be mixed.

FIGURE 42 *Summary of all included studies in the 2004 and 2010 reviews- rivastigmine v. placebo*



4.6.4. Memantine v. placebo

4.6.4.1. Differing approaches to pooling data

A key difference to our memantine findings and Lundbeck's lies in our differing approaches to pooling data. In particular Lundbeck have pooled together memantine + AChEI v. placebo + AChEI trials with memantine monotherapy v. placebo studies; we were not comfortable with this approach due to the heterogeneity of the data. Nevertheless, we have followed this approach for completeness and present the results in Appendix 14. The effect of pooling data in this way is to show a more favourable response to memantine.

4.6.4.2. Identified evidence

The 2004 review lists two RCTs as investigations into the effectiveness of memantine in AD. However, one of those studies – Tariot and colleagues (2004)[135] – addressed the effectiveness of memantine in combination with donepezil; accordingly, this study is considered as part of our assessment of combination therapy (see ¶*4.7.1.3*, *below*). The remaining RCT is that by Reisberg and colleagues (2003).[136]

We identified one additional RCT of relevance to this comparison, details of which are presented in *Table 25*.[137] This studies' interventions, comparators and baseline characteristics can be seen in *Table 26* and markers of internal validity in *Table 27*.

TABLE 25 *Design of included studies – memantine* v. *placebo*

Study details	Inclusion criteria	Exclusion criteria	Methodological notes	Other
Van Dyck et al. (2007)[137] **Design:** Parallel double-blind RCT **Country:** USA **No. of centres:** 35 **No. randomised:** 350 **Maximum follow-up:** 24 **MMSE range included:** 5–14	Probable AD (NINCDS-ADRDA criteria) Age >=50yr Brain imaging evaluation (CT or MRI performed within 12 months before study entry) consistent with probable AD A knowledgable and reliable caregiver to accompany the participant to all study visits and supervise administration of the study drug Ability to ambulate Sufficient vision and hearing to comply with	Significant and active pulmonary, gastrointestinal, renal, hepatic, endocrine, or cardiovascular disease Clinically significant B12 or folate deficiency Evidence of any psychiatric or neurologic disorder other than AD Hachinski Ischaemia Score >4 Delusions or delirium	**Sample attrition / dropout:** 260 of 350 completed study. 90 withdrew after allocation: adverse events (n=45), consent withdrawn (n=26), protocol violation (n=8), insufficient therapeutic response (n=3), other (n=8). No differences between groups. **Randomisation and allocation:** Randomisation procedure not reported **Power calculation:** Assuming an effect size of 0.35, at least 340 participants were needed to provide 90% power at an alpha-level of 0.05 (2-sided) on the basis of	**Therapy common to all participants:** 1 to 2wk single-blind placebo lead-in phase to assess compliance and minimise treatment response at baseline **Study Funding:** Forest Laboratories, Inc provided all financial and material support for the study, as well as statistical and editorial support for the manuscript. **Other conflicts:** Lead author (CD) and 2 co-authors (PT, BM) have received grant support and honoraria from Forest

Study details	Inclusion criteria	Exclusion criteria	Methodological notes	Other
	assessments Medical stability Stable doses of the following medications were allowed: antihypertensives, anti-inflammatories, diuretics, laxatives, antidepressants, atypical antipsychotics, tocopherol	(DSM-IV criteria) Active malignancy History of subnstance abuse within 10yr Likelihood of nursing home placement within 6mo Previous memantine treatment Treatment with an investigational drug within 30dy (or 5 drug half-lives, whichever was longer) of screening Postmenopausal >2yr, or surgically sterile (female participants)	a 2 sample t test for change from baseline to week 24 in SIB and ADCS-ADL scores.	Laboratories, Inc. One co-author (PT) has given expert testimony related to memantine. One author (EM) is an employee of Forest Laboratories, Inc.

TABLE 26 *Interventions, comparators, and baseline characteristics of participants in included studies – memantine* v. *placebo*

Study	Arm	Dose (mg/d)	Dosage details	N	Age	Sex (n male)	Race (n white)	Weight (kg)	Education (yrs)	Duration of dementia (mo)	ADAS-cog	MMSE
Van Dyck et al. (2007)[137]	Memantine	5–20	Initial dosage of 5mg/d with titration in 5mg weekly increments to a final dosage of 20mg/d (administered as two 5mg tablets twice a day). Dose adjustments were permitted between wks 3 and 8 for participants with AEs. Participants unable to tolerate 20mg/dy by the end of week 8 were discontinued from the study. Compliance monitored by inventory of returned individual blister packs, and protocol adherence by routine assessment of concomitant medication use.	178	78.1 (SD 8.20)	49 (27.5%)	142 (79.8%)	64.4 (SD 13.5)[a]				10.0 (SD 2.80)
	Placebo	-	-	172	78.3 (SD 7.60)	51 (29.7%)	141 (82.0%)	65.8 (SD 12.8)				10.3 (SD 3.10)

[a] *n*=176

TABLE 27 *Markers of internal validity of included studies – memantine v. placebo*

	Was the assignment to the treatment groups really random?	Was the treatment allocation concealed?	Were the groups similar at baseline in terms of prognostic factors?	Were the eligibility criteria specified?	Were outcome assessors blinded to the treatment allocation?	Was the care provider blinded?	Was the patient blinded?	Were the point estimates and measure of variability presented for the primary outcome measure?	Did the analyses include an intention-to-treat analysis?	Were withdrawals and dropouts completely described?
Van Dyck *et al.* (2007)[137]	UNKNOWN	UNKNOWN	REPORTED – YES	YES	UNKNOWN	PARTIAL	PARTIAL	ADEQUATE	ADEQUATE	ADEQUATE

4.6.4.3. Evidence of clinical effectiveness

4.6.4.3.1. Cognition

The 2004 report by Loveman and colleagues found two studies of memantine; they summarised their results as follows:

> *"Both studies used the SIB as a measure of cognitive outcome. Statistically significant differences in favour of the use of memantine over placebo were apparent in the two studies. MMSE scores deteriorated in both the memantine group and the placebo group and the degree of deterioration was not statistically significantly different between the two groups."*[2]

New data

The data from the new trial only showed a significant effect from memantine on one of six analyses. However, this was in an observed cases only analysis which may have biased the results; *Table 28* summarises the results.

TABLE 28 *Measures of cognition in included studies – memantine* v. *placebo*

Study	Subgroup	Outcome	Type	Memantine		Placebo		*p*
				N	Mean	N	Mean	
Van Dyck et al. (2007)[137]	LOCF analysis	SIB – 24wk	Mean change	170	-2 (SD 13)	165	-2.5 (SD 12.8)	0.616[a]
	OC population	SIB – 4wk	Mean change	167	0.875 (SD 7.43)[b]	164	-0.3 (SD 6.4)[b]	0.146[a]
		SIB – 8wk	Mean change	158	2.08 (SD 7.86)[b]	155	0.375 (SD 7.16)[b]	0.064[a]
		SIB – 12wk	Mean change	146	1.65 (SD 9.06)[b]	150	-0.825 (SD 8.27)[b]	0.008[a]
		SIB – 18wk	Mean change	140	0 (SD 8.28)[b]	139	-2.12 (SD 9.14)[b]	0.065[a]
		SIB – 24wk	Mean change	131	-1.8 (SD 12.6)	126	-2.4 (SD 13.5)	0.617[a]

[a] ANCOVA (treatment group and centre as main effects; baseline score as covariate)

[b] estimated from figure

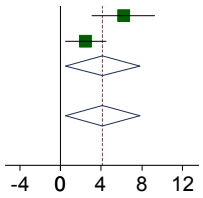
Synthesis with existing evidence-base

Severe Impairment Battery

The data from the new trial was pooled with that of the existing studies in random-effects meta-analyses of the Severe Impairment Battery (SIB) at 12 weeks and 24-28 weeks. The results showed a significant effect at 12 weeks, (WMD=4.15 (95%CI 0.52, 7.78), p=0.025) but not at 24–28 weeks (*Figure 43* and *Figure 44)*.

FIGURE 43 *Random-effects meta-analysis: SIB at 12wk (mean change from baseline) – memantine v. placebo*

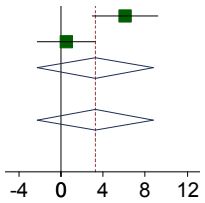
	Memantine			Placebo				
	N	mean	SD	N	mean	SD	WMD (95%CI)	Wght
OC population								
Reisberg et al. (2003)[136]	107	0.80	10.34	106	-5.40	12.35	6.200 (3.138, 9.262)	44.9
Van Dyck et al. (2007)[137]	146	1.65	9.06	150	-0.83	8.27	2.475 (0.497, 4.453)	55.1
subtotal (Q=4.01 [p on 1 d.f.=0.045]; I^2=75.1%; τ^2=5.209)							**4.147 (0.515, 7.778)**	**100.0**
							p=0.025	
Overall pooled estimate							**4.147 (0.515, 7.778)**	
(Q=4.01 [p on 1 d.f.=0.045]; I^2=75.1%; τ^2=5.209)							p=0.025	
Small-study effects: not calculable								



favours placebo ***favours memantine***

FIGURE 44 *Random-effects meta-analysis: SIB at 24–28wk (mean change from baseline) – memantine v. placebo*

	Memantine			Placebo				
	N	mean	SD	N	mean	SD	WMD (95%CI)	Wght
LOCF analysis								
Reisberg et al. (2003)[136]	124	-4.00	11.34	123	-10.10	13.50	6.100 (2.989, 9.211)	49.2
Van Dyck et al. (2007)[137]	170	-2.00	13.04	165	-2.50	12.85	0.500 (-2.272, 3.272)	50.8
subtotal (Q=6.94 [p on 1 d.f.=0.008]; I^2=85.6%; τ^2=13.421)							**3.254 (-2.233, 8.741)**	**100.0**
							p=0.245	
Overall pooled estimate							**3.254 (-2.233, 8.741)**	
(Q=6.94 [p on 1 d.f.=0.008]; I^2=85.6%; τ^2=13.421)							p=0.245	
Small-study effects: not calculable								



favours placebo ***favours memantine***

4.6.4.3.2. Functional

The previous assessment report summarized the findings about the effects of memantine on functional outcomes as:

> *"Both studies demonstrated that memantine appears to show a statistically significant benefit to participants on the ADCS-ADL when compared to placebo, with a reduction in the level of deterioration."*[2]

New data

The results from the new study showed no significant benefit on functional outcomes for memantine compared to placebo, (*Table 29*).

TABLE 29 *Measures of functional ability in included studies – memantine* v. *placebo*

Study	Subgroup	Outcome	Type	Memantine		Placebo		
				N	Mean	N	Mean	*p*
Van Dyck et al. (2007)[137]	LOCF analysis	ADCS-ADL – 24wk	Mean change	171	-2 (SD 7.85)	165	-2.7 (SD 7.71)	0.282[a]
		FAST – 24wk	Mean change	151	0.3 (SD 1.23)	141	0.6 (SD 1.19)	0.093[a]
	OC population	ADCS-ADL – 4wk	Mean change	168	0.312 (SD 4.37)[b]	164	0.512 (SD 4)[b]	0.801[a]
		ADCS-ADL – 8wk	Mean change	159	-0.0875 (SD 5.2)[b]	156	-0.188 (SD 4.84)[b]	0.665[a]
		ADCS-ADL – 12wk	Mean change	147	0 (SD 5.46)[b]	150	-0.488 (SD 5.05)[b]	0.155[a]
		ADCS-ADL – 18wk	Mean change	142	-0.688 (SD 7.3)[b]	140	-1.38 (SD 5.62)[b]	0.357[a]
		ADCS-ADL – 24wk	Mean change	133	-1.3 (SD 6.92)	127	-2.3 (SD 6.76)	0.188[a]
		FAST – 24wk	Mean change	133	0.3 (SD 1.15)	127	0.6 (SD 1.13)	0.074[a]

[a] ANCOVA (treatment group and centre as main effects; baseline score as covariate)

[b] estimated from figure

Synthesis with existing evidence-base

The data from the new studies were synthesized with the existing evidence in random-effects meta-analysis.

Alzheimer's Disease Cooperative Study Activities of Daily Living Inventory

Two studies provide data on functional effect as measured by ADCS-ADL; both report the modified ADCS-ADL$_{19}$ version of the instrument, consisting of 19 items that have been individually validated in cases of more severe dementia. The data were meta-analysed at 12 weeks and 24–28 weeks follow-up. The results were not significant at 12 weeks and barely significant at 24–28 weeks, especially considering that the population analysed were LOCF, WMD=1.41 (95%CI 0.04, 2.78), p=0.044 (*Figure 45* and *Figure 46*).

FIGURE 45 *Random-effects meta-analysis: ADCS-ADL$_{19}$ at 12wk (mean change from baseline) – memantine v. placebo*

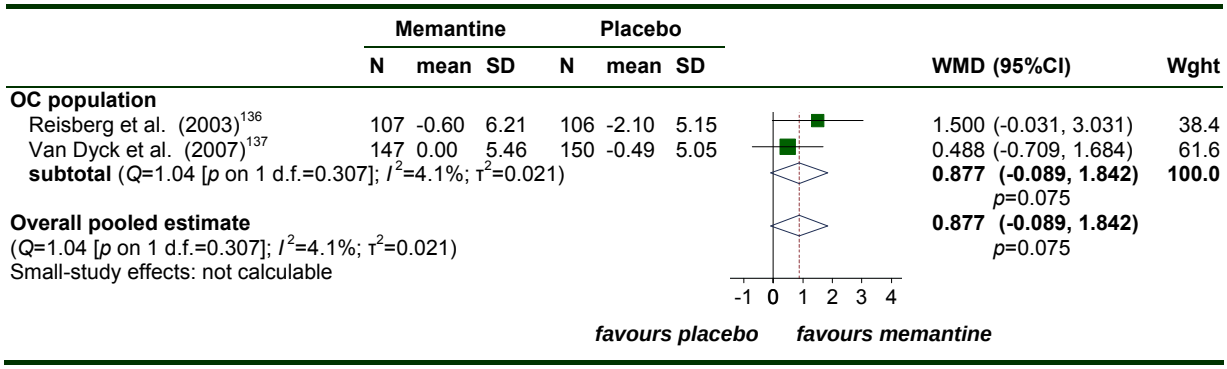
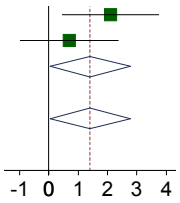


FIGURE 46 *Random-effects meta-analysis: ADCS-ADL$_{19}$ at 24–28wk (mean change from baseline) – memantine v. placebo*

	Memantine			Placebo				
	N	mean	SD	N	mean	SD	WMD (95%CI)	Wght
LOCF analysis								
Reisberg et al. (2003)[136]	124	-3.10	6.79	123	-5.20	6.33	2.100 (0.463, 3.737)	50.6
Van Dyck et al. (2007)[137]	171	-2.00	7.85	165	-2.70	7.71	0.700 (-0.963, 2.363)	49.4
subtotal (Q=1.38 [p on 1 d.f.=0.240]; I^2=27.7%; τ^2=0.271)							**1.408 (0.036, 2.780)**	**100.0**
							p=0.044	
Overall pooled estimate							**1.408 (0.036, 2.780)**	
(Q=1.38 [p on 1 d.f.=0.240]; I^2=27.7%; τ^2=0.271)							p=0.044	
Small-study effects: not calculable								

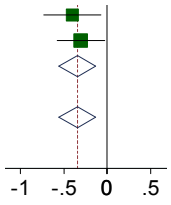


Functional Assessment Staging Tool

Another meta-analysis was conducted with data from existing and new studies using the Functional Assessment Staging Tool (FAST) at 24–28 weeks follow-up. The overall pooled estimate showed a significant benefit from memantine compared to placebo, WMD=-0.34 (95%CI -0.55, -0.13), p=0.002 (*Figure 47*).

FIGURE 47 *Random-effects meta-analysis: FAST at 24–28wk (mean change from baseline) – memantine v. placebo*

	Memantine			Placebo				
	N	mean	SD	N	mean	SD	WMD (95%CI)	Wght
LOCF analysis								
Reisberg et al. (2003)[136]	121	0.20	1.24	118	0.60	1.39	-0.400 (-0.734, -0.066)	40.7
Van Dyck et al. (2007)[137]	151	0.30	1.23	141	0.60	1.19	-0.300 (-0.577, -0.023)	59.3
subtotal (Q=0.2 [p on 1 d.f.=0.652]; I^2=0.0%; τ^2=0.000)							**-0.341 (-0.554, -0.127)**	**100.0**
							p=0.002	
Overall pooled estimate							**-0.341 (-0.554, -0.127)**	
(Q=0.2 [p on 1 d.f.=0.652]; I^2=0.0%; τ^2=0.000)							p=0.002	
Small-study effects: not calculable								



4.6.4.3.3. Behavioural and mood

The 2004 assessment report summarised the finding for behavioural outcomes comparing memantine with placebo as:

> *"It appears that participants receiving memantine and already receiving a steady dose of donepezil have a statistically significantly lower NPI score than placebo. Those on memantine only however, showed no statistically significant difference compared to placebo."*[2]

New data

The study that was published after 2004 measured behavioural outcomes using the NPI and the Behavioural Rating Scale for Geriatric Patients (BGP). Neither measure showed a significant benefit from memantine (*Table 30*).

TABLE 30 *Measures of behavioural effect and mood in included studies – memantine* v. *placebo*

Study	Subgroup	Outcome	Type[a]	Memantine		Placebo		*p*
				N	Mean	N	Mean	
van Dyck et al. (2007)[137]	LOCF analysis	NPI – 24wk	MC	161	1 (SD 16.5)	154	1.1 (SD 17.4)	0.963[b]
		BGP: total – 24wk	MC	151	0.6 (SD 6.14)	141	1.5 (SD 7.12)	0.197[b]
		BGP: care dependency – 24wk	MC	151	0.5 (SD 4.92)	141	1.4 (SD 4.75)	0.076[b]
	OC population	NPI – 24wk	MC	133	0.5 (SD 15)	127	1 (SD 15.8)	0.782[b]
		BGP: total – 24wk	MC	133	0.4 (SD 6.92)	127	1.1 (SD 6.76)	0.312[b]
		BGP: care dependency – 24wk	MC	133	0.4 (SD 4.61)	127	1.2 (SD 5.63)	0.138[b]

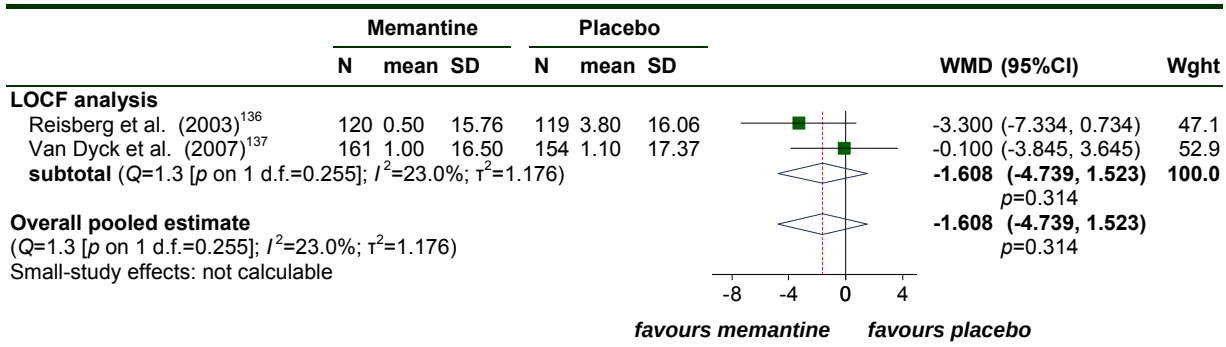
[A] MC=Mean change

[b] ANCOVA (treatment group and centre as main effects; baseline score as covariate)

Synthesis with existing evidence-base

The NPI data from van Dyck and colleagues was pooled with the existing data at 24–28 weeks follow-up in a random-effects meta-analysis. This analysis also failed to show a significant gain from memantine compared to placebo *(Figure 48*).

FIGURE 48 *Random-effects meta-analysis: NPI at 24–28wk (mean change from baseline) – memantine v. placebo*



4.6.4.3.4. Global effect

In 2004 Loveman and colleagues summarised the results for global measures comparing memantine with placebo as:

> *"both studies used the CIBIC-Plus as a measure of global outcome, and in both cases, memantine appeared to be effective."*[2]

New data

van Dyck and colleagues also measured global outcomes with the CIBIC-plus; however, the differences they found were not significant, (*Table 31*).

TABLE 31 *Measures of global effect in included studies – memantine* v. *placebo*

Study	Subgroup	Outcome	DataType	Memantine		Placebo		*p*
				N	Mean	N	Mean	
Van Dyck et al. (2007)[137]	LOCF analysis	CIBIC-plus score - 24wk	Continuous	171	4.3 (SD 13.1)	163	4.6 (SD 12.8)	0.182a
	OC population	CIBIC-plus score - 24wk	Continuous	134	4.3 (SD 12.7)	127	4.6 (SD 11.3)	0.089a

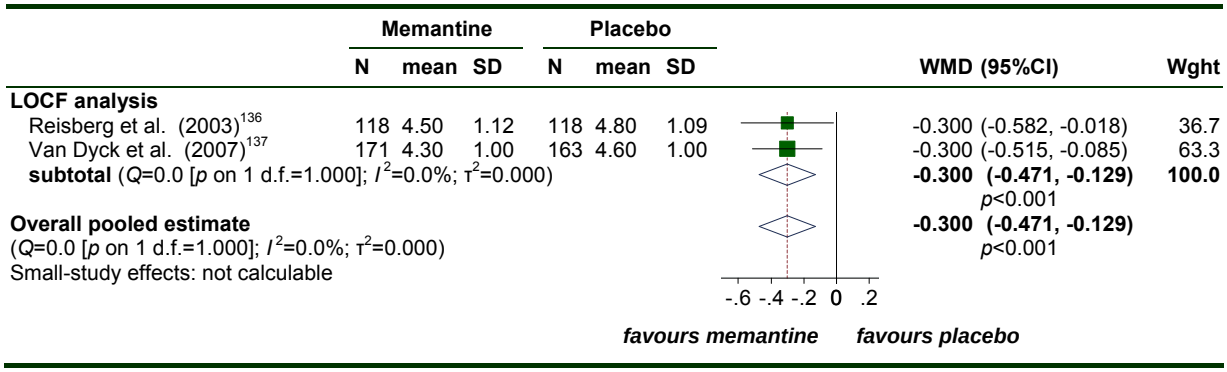
[a] Cochran-Mantel-Haenszel statistic using modified Ridit scores (Van Elteren test) controlling for study centre

Synthesis with existing evidence-base

Clinician Interview-Based Impression of Change

When the new data were pooled with the existing studies in a random-effects meta-analysis the overall pooled estimate showed a significant beneficial effect from memantine compared to placebo, WMD=-0.30 (95%CI -0.47, -0.13), p<0.001 (*Figure 49*).

FIGURE 49 *Random-effects meta-analysis: CIBIC-plus at 24–28wk – memantine v. placebo*



4.6.4.3.5. Quality of life

None of the included studies provided any randomised evidence on QoL with memantine compared with placebo, and no such data were identified in the 2004 review.

4.6.4.3.6. Safety

The proportion of any AEs were similar in treatment and control groups (T=74%m C=73%). The main AEs in the memantine group were agitation and hypertension, and agitation and falls in the control group (*Table 32*).

TABLE 32 *AEs in included studies – memantine* v. *placebo*

Adverse event	Van Dyck et al. (2007)[137]				
	Memantine		Placebo		*p*
	N	n (%)	N	n (%)	
Any AE	178	131 (73.6%)	172	125 (72.7%)	0.941[a]
Any serious AE	178	26 (14.6%)	172	29 (16.9%)	0.666[a]
Diarrhoea	178	10 (5.6%)	172	8 (4.7%)	0.867[a]
Agitation	178	16 (9.0%)	172	24 (14.0%)	0.197[a]
Anxiety	178	10 (5.6%)	172	6 (3.5%)	0.485[a]
Depression	178	9 (5.1%)	172	5 (2.9%)	0.451[a]
Injury	178	10 (5.6%)	172	13 (7.6%)	0.605[a]
Dizziness	178	12 (6.7%)	172	11 (6.4%)	0.932[a]
Headache	178	3 (1.7%)	172	11 (6.4%)	0.048[a]
Urinary tract infection	178	9 (5.1%)	172	9 (5.2%)	0.867[a]
Fall	178	10 (5.6%)	172	17 (9.9%)	0.195[a]
Influenza-like symptoms	178	10 (5.6%)	172	8 (4.7%)	0.867[a]
Confusion	178	9 (5.1%)	172	8 (4.7%)	0.942[a]
Hypertension	178	14 (7.9%)	172	4 (2.3%)	0.035[a]
Peripheral oedema	178	12 (6.7%)	172	8 (4.7%)	0.541[a]
Constipation	178	11 (6.2%)	172	8 (4.7%)	0.693[a]
Insomnia	178	4 (2.2%)	172	9 (5.2%)	0.233[a]

[a] chi-square test (Yates's correction) (calculated by reviewer)

4.6.4.4. Summary: memantine v. placebo

One new moderate-to-poor quality study was found to add to the existing evidence for memantine v. placebo.

The pooled results for cognitive abilities measured by the SIB, showed a significant benefit from memantine at 12 weeks follow-up (WMD=4.15 (95%CI 0.52, 7.78) p=0.025). However, at 24 weeks the data pooled with that of the previous review showed no significant benefit.

Similar to the previous review, the new study found a significant benefit from memantine from the FAST functional outcome, although not with the ADCS-ADL at 12 weeks. When the FAST data from new and existing studies were pooled a significant relationship was found between memantine and an improvement in scores, WMD=-0.34 (95%CI -0.55, -0.13), p=0.002. A marginally significant benefit was seen from memantine when pooled ADCS-ADL data were measured at 24–28 weeks, WMD=1.41 (95%CI 0.04, 2.78) p=0.044.

The results from behavioural outcomes in the new study, similar to the previous review, failed to show a significant benefit from memantine, either singly or when the data were pooled.

Although the results of the CIBIC-plus failed to show a significant gain from memantine, when this data was pooled with that from the previous review a significant effect was found, WMD= -3.00 (95% CI-0.471,-0.129), p<0.001.

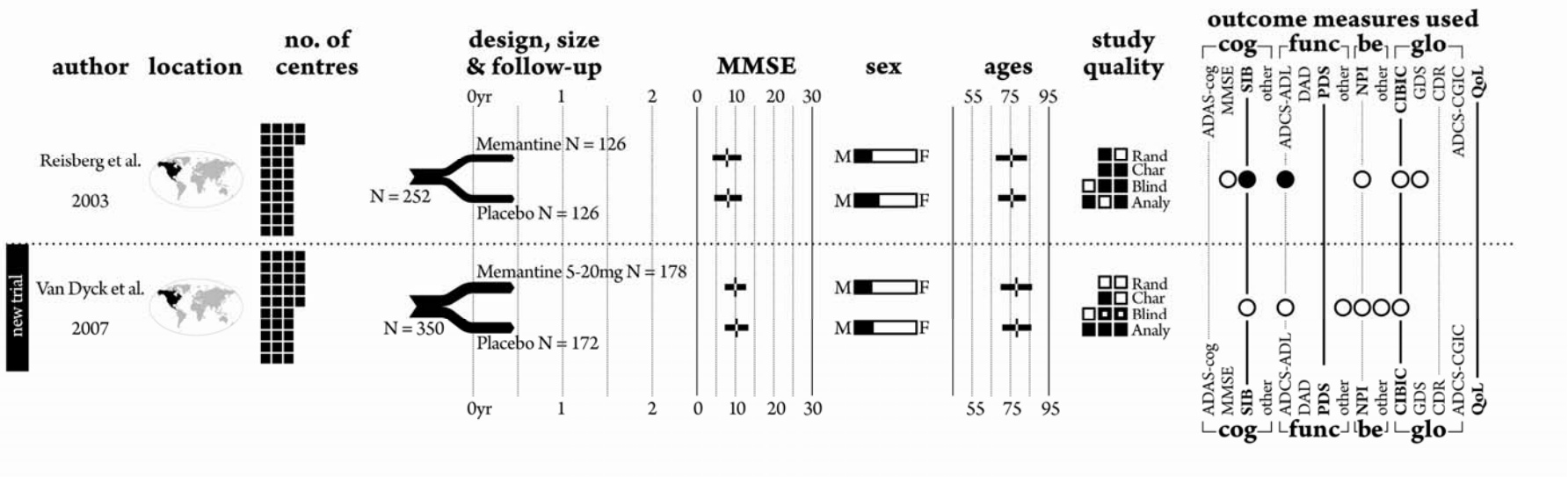
No studies reported QoL outcomes. The main AEs were agitation and hypertension.

The meta-analysis of memantine v. placebo studies showed benefit from memantine at 12 weeks follow-up on the SIB. However, treatment gain, measured by functional outcome, depended on the type of instrument used, and no benefit was seen from behavioural outcomes. Nevertheless, pooled estimates of global outcomes showed a benefit from taking memantine at 24–28 weeks. Overall the pooled results from these moderate to poor quality studies showed inconclusive results for cognitive and behavioural outcomes. The results for functional outcomes were dependent on the measure used but the pooled results of new and existing evidence for global outcomes showed significant benefit from using memantine.

4.6.4.5. Graphical summary of memantine v. placebo

Figure 50 below, illustrates how little the evidence has changed for memantine v. placebo. The cognitive benefits found in 2004 failed to be replicated; indeed the new study only favoured memantine on one outcome measure. However, the quality of the new and existing studies was not high; thus, these results cannot be considered conclusive.

FIGURE 50 *Summary of all included studies in the 2004 and 2010 reviews- memantine v. placebo*



4.7. Head-to-head comparisons

4.7.1.1. Identified evidence

Alongside placebo-controlled trials, a certain amount of randomised evidence provides direct, head-to-head comparisons of two or more of the technologies under review. Three such RCTs were included in the 2004 review: Fuschillo and colleagues. (2001)[138] (donepezil *v.* rivastigmine), Wilkinson and colleagues. (2002)[139] (donepezil *v.* rivastigmine), and Jones and colleagues. (2004)[140] (donepezil *v.* galantamine).

Our searches identified a further four RCTs of this type. Details of the design of these trials are tabulated in *Table 33,* and a summary of treatments and baseline characteristics of participants can be found in *Table 34*. Two of the new RCTs compared all three AChEIs (although Nordberg and colleagues' trial[141] is only of relevance to the current review for its safety data). One trial investigated donepezil v. rivastigmine,[142] and the last was concerned with donepezil v. galantamine.[143]

TABLE 33 *Design of included studies – head-to-head comparisons*

Study details	Inclusion criteria	Exclusion criteria	Methodological notes	Other
Ancoli-Israel et al. (2005)[143] **Design:** Parallel double-blind RCT **Country:** Not reported. All study authors based in USA **No. of centres:** Not reported **No. randomised:** 63 **Maximum follow-up:** 8 **MMSE range included:** 10–24	Mild to moderate AD (criteria not reported) age ≥60 Resident with a responsible caregiver who agreed to participate and monitor sleep and answer questionnaires	Other neurodegenerative disease contributing to dementia (including mulit-infarct dementia or clinically active cerebrovascular disease) Other medical conditttions causing cognitive impairment Clinically significant co-existing medical conditions Use of a muscarinic-1 agonist or AChEI within 30d prior to involvement	**Sample attrition / dropout:** 54 of 63 completed study; discontinued due to adverse event (n=3 in galantamine arm; n=4 in donepezil arm); discontinued due to severe adverse event possibly related to trial drug (hepatic failure, n=1 in donepezil arm); death (judged to be unrelated to trial drug, n=1) **Randomisation and allocation:** Randomisation procedure not described **Power calculation:** None	**Therapy common to all participants:** 2-week, single-blind, placebo run-in **Study Funding:** Janssen Medical Affairs **Other conflicts:** Lead author declares no financial disclosure; co-authors are employees of funder (Janssen Medical Affairs)
Bullock et al. (2005)[142] **Design:** Parallel double-blind RCT **Country:** Australia, Canada, France, Germany, Italy,	Male or female outpatients aged 50-85yrs AD (DSM-IV criteria) or probable AD (NINCDS-ADRDA criteria) Contact with a responsible	Current diagnosis of any primary neurodegenerative disorder other than AD (including Parkinson's disease) Any advance, severe, progressive or unstable disease or disability	**Sample attrition / dropout:** 578 of 994 (58.1%) completed study (rivastigmine 261 of 495 (52.7%), donezepil 317 of 499 (63.5%) (998 were randomised, 4 withdrew before receiving treatment)	**Therapy common to all participants:** None **Study Funding:** Study supported by Novartis Pharma AG 4 of the study authors (YH, JN, GR, RL) are employees of Novartis The remaining 4 authors

Study details	Inclusion criteria	Exclusion criteria	Methodological notes	Other
Spain, UK **No. of centres:** 94 **No. randomised:** 998 **Maximum follow-up:** 104 **MMSE range included:** 10–20	caregiver at least once a day (Patients with AD who also had symptoms suggestive of concomitant Lewy body disease (McKeith et al criteria) were also permitted to enter the study	A major depressive episode Active, uncontrolled seizure disorder or peptic ulceration Acute, severe or unstable asthmatic conditions Severe or unstable cardiovascular disease History or diagnosis of cerebrovascular disease Known hyperensitivity to drugs similar to rivastigmine or donezepil in structure or pharmacologic action Use of any cholinesterase inhibitor or other approved treatment for AD in the 6 weeks prior to randomisation Use of any investigational drug, any drug or treatment known to cause major organ system toxicity, or any new psychotropic medication during the 4 weeks prior to randomisation Anticholinergic drugs at randomisation	Reasons for non-completion: rivastigmine – adverse events (n=129); abnormal lab values (n=1); unsatisfactory therapeutic effect (n=19); protocol violation (n=12); withdrawn consent (n=34); lost to follow-up (n=10); admiistrative problems (n=4); death (n=26) donezepil – adverse events (n=80); abnormal lab values (n=1); unsatisfactory therapeutic effect (n=17); protocol violation (n=9); withdrawn consent (n=22); lost to follow-up (n=13); admiistrative problems (n=6); death (n=34) **Randomisation and allocation:** Performed using Interactive Voice Response System that automated the random assignment of treatment groups to randomisation numbers. Randomisation was stratified with respect to severity, i.e. was done separately with MMSE scores of 10-14 and 15-20. All treatments were supplied as capsules that were identical in size, shape and colour, and all patients received the same number of capsules per day. **Power calculation:** Powered at 85% to detect a statistically significant (significance level 5%, two-sided) difference in SIB of 4 points between the two groups (assuming a SD of 20 on change from baseline in mean SIB scores, as observed in previous trials), sample size of 450 patients per treatment group was required.	(RB, JT, HB, GG) did not receive remuneration for taking part in the study or writing the manuscript **Other conflicts:** Study supported by Novartis Pharma AG 4 of the study authors (YH, JN, GR, RL) are employees of Novartis The remaining 4 authors (RB, JT, HB, GG) did not receive remuneration for taking part in the study or writing the manuscript
Cumbo (2005)[92] **Design:** - **Country:** Funded by an Italian health agency, but not stated whether study conducted in Italy or elsewhere. **No. of centres:**	Probable AD (NINCS-ARDRA) >=3yr duration of disease No behavioural symptoms Carer who could ensure compliance to treatment and attendance and provide the	History of primary neurological or psychiatric disease other than AD Drug or alcohol abuse Clinically significant medical or surgical disorders independently of stability Previous therapy for dementia	**Sample attrition / dropout:** None **Randomisation and allocation:** No details of randomisation procedure reported. Open-label trial. **Power calculation:** None reported	**Therapy common to all participants:** None **Study Funding:** Supported by Department of Neuroscience (NHS District of Caltanissetta) Novartis Farma SpA supported the English editing of the manuscript **Other conflicts:** Supported by Department of

Study details	Inclusion criteria	Exclusion criteria	Methodological notes	Other
Not stated. Small sample size suggests single centre. **No. randomised:** 101 **Maximum follow-up:** 78 **MMSE range included:** 10–27	information required for psychometric and behavioural assessments	Concomitant treatment with cholinomimetic or anticholinergic drugs, investigational drugs, tricyclic antidepressants or neuroleptics Refusal to give informed consent in writing		Neuroscience (NHS District of Caltanissetta) Novartis Farma SpA supported the English editing of the manuscript
Nordberg et al. (2009)[141] **Design:** - **Country:** Not reported **No. of centres:** Not reported **No. randomised:** 63 **Maximum follow-up:** 13 **MMSE range included:** 10–20	AD (DSM-IV criteria) and probable or possible AD (NINCDS-ADRDA criteria) Age 50-85yr Provided the dose had been stabilised for the past month, treatment with psychotropics was permitted	Prior exposure to rivastigmine, donepezil or galantamine Advance, severe or unstable disease of any type that might interfere with study evaluation or put the patient at special risk Imaging findings consistent with a condition other than AD that would explain the patient's dementia Current treatment with coumarin derivatives Blood clotting abnormalities or inadequate platelet function	**Sample attrition / dropout:** 53 of 63 completed study. 10 withdrew after allocation; adverse events (n=8), withdrew consent (n=1), lost to follow-up (n=1) **Randomisation and allocation:** Randomisation procedure not described. Open-label trial (although laboratory personnel who processed CSF samples were blinded). **Power calculation:** Assuming a mean treatment difference of 0.3 U/L (primary outcome variable), SD 0.28 and two-sided significance level of 0.025, z-test showed approximately 20 patients per treatment group were required to achieve a power of 0.85 for detecting a significant pairwise treatment difference.	**Therapy common to all participants:** None **Study Funding:** Novartis Pharmaceuticals; Swedish Research Council; KI foundations, L-H Osterman and Stohne's Foundations supported two co-authors (AN, TDS). Alpha-Plus provided editorial assistance with the production of the manuscript. **Other conflicts:** Three co-authors (AN, TD-S, MM) were responsible for the enzyme analysis and received research sponsorship from Novartis. One co-author's (HS) institute received research sponsorship from Novartis for this study. Two co-authors (GE, RL) are fulltime employees of Novartis.

TABLE 34 *Interventions, comparators, and baseline characteristics of participants in included studies – head-to-head comparisons*

Study	Arm	Dose (mg/d)	Dosage details	N	Age	Sex (n male)	Race (n white)	Weight (kg)	Education (yrs)	Duration of dementia (mo)	ADAS-cog	MMSE
Bullock et al. (2005)[142]	Rivastigmine	3–12	Titrated from 3mg/d for the first 4wk up to a maximum of 12mg/d in increments of 3mg/d every 4wk	495	75.9 (SD 6.60)	154 (31.1%)				33.6 (SD 22.2)		15.1 (SD 3.00)
	Donepezil	5–10	Titrated from 5mg/d for wks 1–8up to 10mg/d in wks 9–16 For patients who did not achieve the maximum dose during the titration period, investigators were asked to make at least one attempt during the maintenance period to increase the dose to the next highest dose level.	499	75.8 (SD 6.80)	157 (31.5%)				34.2 (SD 26.5)		15.1 (SD 2.90)
Cumbo (2005)[92]	Rivastigmine	9[a]	dosage / titration scheme not reported	101[b]	76.4 (rng 66–83)[b]	43 (42.6%)[b]			5.00 (rng 3–12)[b]	61.1 (rng 36–108)[b]		16.6[b]
	Galantamine	16[a]	dosage / titration scheme not reported									
	Donepezil	10[a]	dosage / titration scheme not reported									
Ancoli-Israel et al. (2005)[143]	Donepezil	5–10	Dose titrated from 5mg once a day at night for the first 4wk up to 10mg once a day at night for remainder of study	32	77.8 (SD 6.20)	14 (43.8%)	26 (81.3%)		[c]			19.4 (rng 13–24)
	Galantamine	8–16	Dose titrated from 4mg twice a day for the first 4wk up to 8mg twice a day for remainder of study	31	76.5 (SD 7.70)	10 (32.3%)	25 (80.6%)		[d]			19.3 (rng 11–24)
Nordberg et al. (2009)[141]	Donepezil	5–10	starting dose 5mg qd; after >=4wk, if tolerated, up-titrated to 10mg qd; no subsequent up-titrations	20	74.0 (SD 8.00)	9 (45.0%)	20 (100.0%)	65.2 (SD 8.00)		32.4 (SD 19.2)		20.0 (SD 3.50)
	Galantamine	8–24	starting dose 4mg bd; after >=4wk, if tolerated, up-titrated to 8mg bd; subsequent up-titrations could be made after >=4wk at each dose, based upon the patient's well-being and tolerability, to a maximum of 12mg bd	21	73.7 (SD 6.50)	5 (23.8%)	21 (100.0%)	65.7 (SD 11.5)		39.6 (SD 25.2)		19.2 (SD 3.10)
	Rivastigmine	3–12	starting dose 1.5mg bd; after >=4wk, if tolerated, up-titrated to 3mg bid; subsequent up-titrations could be made after >=4wk at each dose, based upon the patient's well-being and tolerability, to a maximum of 6mg bid	22	76.8 (SD 8.90)	5 (22.7%)	21 (95.5%)	65.1 (SD 9.70)		34.8 (SD 25.2)		18.8 (SD 3.80)

[a] mean dose received during trial; allowable regimen not described

[b] whole trial population; no data presented for individual arms

[c] 26 (81.3%) at least high-school

[d] 22 (71.0%) at least high-school

TABLE 35 *Markers of internal validity of included studies – head-to-head comparisons*

	Was the assignment to the treatment groups really random?	Was the treatment allocation concealed?	Were the groups similar at baseline in terms of prognostic factors?	Were the eligibility criteria specified?	Were outcome assessors blinded to the treatment allocation?	Was the care provider blinded?	Was the patient blinded?	Were the point estimates and measure of variability presented for the primary outcome measure?	Did the analyses include an intention-to-treat analysis?	Were withdrawals and dropouts completely described?
Bullock *et al.* (2005)[142]	ADEQUATE	ADEQUATE	REPORTED – YES	INADEQUATE	PARTIAL	ADEQUATE	ADEQUATE	ADEQUATE	ADEQUATE	ADEQUATE
Cumbo (2005)[92]	UNKNOWN	UNKNOWN	UNKNOWN[a]	UNKNOWN	UNKNOWN	UNKNOWN[b]	UNKNOWN[b]	ADEQUATE	ADEQUATE[c]	ADEQUATE[d]
Ancoli-Israel *et al.* (2005)[143]	UNKNOWN	UNKNOWN	REPORTED – YES	UNKNOWN	PARTIAL	PARTIAL	PARTIAL	ADEQUATE	PARTIAL	ADEQUATE
Nordberg *et al.* (2009)[141]	UNKNOWN	UNKNOWN	REPORTED – YES[e]	UNKNOWN	INADEQUATE[f,g]	INADEQUATE	INADEQUATE[f]	ADEQUATE	INADEQUATE	ADEQUATE

[a] Mean or range across all trial arms only given

[b] Open-label trial

[c] All patients completed follow-up

[d] No dropouts occurred

[e] Although note fewer women in donepezil group

[f] Open label trial

[g] Open label trial, monitoring personnel were not blinded (although laboratory personnel who processed CSF samples were blinded)

The quality of newly identified RCTs in this category (*Table 35*) tended to be low. Cumbo's three-way examination of donepezil, galantamine, and rivastigmine[92] is of especially dubious validity, with no description of randomisation or allocation, and an open-label treatment period. Moreover, most of the outcomes reported by this trial – concentrating on distribution of symptoms amongst participants experiencing behavioural disturbance – are of little help for our purpose of establishing the relative effectiveness of the technologies.

Of the RCTs we identified, Bullock and colleagues two-year, double-blind comparison of donepezil and rivastigmine[142] was judged to be much the least susceptible to bias. Robust randomisation, allocation, and assessment methods are reported, and the study was of a good size, with each treatment arm comprising almost 500 individuals.

4.7.1.2. Evidence of clinical effectiveness

4.7.1.2.1. Cognition

New data

Only one of the newly identified RCTs reports outcome measures assessing the cognitive function of participants. Bullock and colleagues[142] report that, following two years of double-blind treatment, a similar cognitive decline was seen in individuals who had been randomised to donepezil or rivastigmine (*Table 36*).

TABLE 36 *Measures of cognition in included studies – head-to-head comparisons*

Study	Subgroup	Outcome	Type	Rivastigmine		Donepezil		*p*
				N	Mean	N	Mean	
Bullock et al. (2005)[142]	LOCF analysis	MMSE – 104wk	Mean change	471	-2.35 (SD 6.51)	484	-2.85 (SD 6.6)	0.089[a] 0.106[b]
		SIB – 104wk	Mean change	471	-9.3 (SD 23.9)	483	-9.91 (SD 24.2)	0.609[a] 0.738[b]

[a] ANCOVA, covarying country, MMSE category, and baseline score

[b] Wilcoxon rank sum test

Synthesis with existing evidence-base

It was not possible to amalgamate new and existing evidence in quantitative synthesis, because Bullock and colleagues' trial[142] featured much more extensive follow-up than the 12–30 week donepezil v. rivastigmine RCTs identified in 2004,[138;139]. Unfortunately, Bullock

and colleagues do not report findings on cognitive measures over the course of their trial (one figure showing SIB decline is provided, but does not give any indication of dispersion at each juncture), so their findings cannot be combined at earlier follow-up, either.

4.7.1.2.2. Functional

New data

Again, Bullock and colleagues' RCT provides the only new evidence on the relative effectiveness of the technologies under review in the functional domain. In the primary – ITT LOCF – analysis, a significant advantage for rivastigmine over donepezil after two years' treatment was detected. Individuals who had been randomised to receive rivastigmine declined by around two fewer points on the ADCS ADL instrument (*Table 37*). It should be noted, however, that this finding was not replicated in secondary analyses, which relied on evaluable cases (all participants who were treated for at least 16 weeks, with LOCF imputation for subsequent missing values) and OCs.

TABLE 37 *Measures of functional ability in included studies – head-to-head comparisons*

Study	Subgroup	Outcome	Type	Rivastigmine		Donepezil		*p*
				N	Mean	N	Mean	
Bullock et al. (2005)[142]	LOCF analysis	ADCS-ADL - 104wk	Mean change	454	-12.8 (SD 19.2)	475	-14.9 (SD 19.6)	0.047[a] 0.007[b]

[a] ANCOVA, covarying country, MMSE category, and baseline score

[b] Wilcoxon rank sum test

Synthesis with existing evidence-base

Again, the much longer duration of the new trial, coupled with its lack of intermediate follow-up data, makes it impossible to perform quantitative synthesis combining Bullock and colleagues' data with that identified in 2004.

4.7.1.2.3. Behavioural and mood

New data

Bullock and colleagues[142] found no significant difference between donepezil and rivastigmine on the NPI scale, with participants in both groups declining by an average of between two and three points over two years' treatment (*Table 38*). Cumbo's trial[92] is explicitly focused on

behavioural disturbance in individuals taking AChEIs. However, the paper mostly concentrates on the profile of individuals who were adjudged to experience behavioural and psychological symptoms, rather than the incidence of such events in the whole population. It was found that most categories of behavioural event happened with lower frequency among those taking rivastigmine; however, no tests of the magnitude of such differences are presented. We have found that, in most cases, any discrepancy would be insufficient to fulfil conventional definitions of statistical significance (i.e. $p<0.05$ by chi-squared test), with the exceptions of the night-time behaviour subdomain of the NPI and the diurnal cycle disturbances item of the Behavioral Pathology in AD (BEHAVE-AD) scale (*Table 38*). The high probability of type I error in the presence of multiple comparisons must clearly be borne in mind, here.

Individuals taking rivastigmine were also reported to have a higher probability of remaining free of behavioural symptoms at 18 months than those taking donepezil in Cumbo's RCT,[92] although the methods adopted in the time-to-event analysis are unclear.

TABLE 38 *Measures of behavioural effect and mood in included studies – head-to-head comparisons*

Study	Subgroup	Outcome	Type[a]	Rivastigmine		Galantamine		Donepezil		*p*		
				N	Mean / n (%)	N	Mean / n (%)	N	Mean / n (%)	RvG	RvD	GvD
Bullock et al. (2005)[142]	LOCF analysis	NPI – 104wk	MC	471	2.4 (SD 17.4)	-	-	484	2.94 (SD 17.6)	-	0.554[b] 0.505[c]	-
Cumbo (2005)[92]	ITT population	Probability of being BPSD-free at 78wk	TTE	37	0.622 (SEM 0.080)	33	0.546 (SEM 0.087)	31	0.484 (SEM 0.090)	0.235[d]	0.055[d]	0.365[d]
		NPI - delusions - 78wk	D	37	1(2.7%)	33	4(12.1%)	31	5(16.1%)	0.288[e]	0.130[e]	0.919[e]
		NPI - hallucinations - 78wk	D	37	0(0.0%)	33	0(0.0%)	31	3(9.7%)	0.341[e]	0.226[e]	0.274[e]
		NPI - agitation/aggression - 78wk	D	37	4(10.8%)	33	9(27.3%)	31	7(22.6%)	0.144[e]	0.326[e]	0.885[e]
		NPI - depression/dysphoria - 78wk	D	37	13(35.1%)	33	10(30.3%)	31	13(41.9%)	0.861[e]	0.746[e]	0.479[e]
		NPI - anxiety - 78wk	D	37	14(37.8%)	33	15(45.5%)	31	14(45.2%)	0.687[e]	0.716[e]	0.820[e]
		NPI - elation/euphoria - 78wk	D	37	0(0.0%)	33	0(0.0%)	31	1(3.2%)	0.341[e]	0.902[e]	0.965[e]
		NPI - apathy/indifference - 78wk	D	37	7(18.9%)	33	7(21.2%)	31	8(25.8%)	0.952[e]	0.698[e]	0.890[e]
		NPI - disinhibition - 78wk	D	37	0(0.0%)	33	3(9.1%)	31	1(3.2%)	0.252[e]	0.902[e]	0.651[e]
		NPI - irritability/lability - 78wk	D	37	12(32.4%)	33	14(42.4%)	31	15(48.4%)	0.538[e]	0.276[e]	0.820[e]
		NPI - aberrant motor behaviour - 78wk	D	37	0(0.0%)	33	0(0.0%)	31	0(0.0%)	0.341[e]	0.355[e]	0.328[e]
		NPI - night-time behaviour - 78wk	D	37	1(2.7%)	33	9(27.3%)	31	0(0.0%)	0.010[e]	0.902[e]	0.008[e]
		NPI - appetite/eating change - 78wk	D	37	0(0.0%)	33	1(3.0%)	31	1(3.2%)	0.936[e]	0.902[e]	0.500[e]
		Developing BPSD - 78wk	D	37	14(37.8%)	33	15(45.5%)	31	16(51.6%)	0.687[e]	0.371[e]	0.808[e]
		BEHAVE-AD - delusional and paranoid ideation - 78wk	D	37	1(2.7%)	33	4(12.1%)	31	5(16.1%)	0.288[e]	0.130[e]	0.919[e]
		BEHAVE-AD - hallucinations - 78wk	D	37	0(0.0%)	33	0(0.0%)	31	3(9.7%)	0.341[e]	0.226[e]	0.274[e]
		BEHAVE-AD - activity disturbances - 78wk	D	37	0(0.0%)	33	0(0.0%)	31	0(0.0%)	0.341[e]	0.355[e]	0.328[e]
		BEHAVE-AD - aggression - 78wk	D	37	4(10.8%)	33	9(27.3%)	31	7(22.6%)	0.144[e]	0.326[e]	0.885[e]
		BEHAVE-AD - diurnal cycle disturbances - 78wk	D	37	1(2.7%)	33	9(27.3%)	31	10(32.3%)	0.010[e]	0.003[e]	0.871[e]
		BEHAVE-AD - affective disturbances - 78wk	D	37	13(35.1%)	33	10(30.3%)	31	13(41.9%)	0.861[e]	0.746[e]	0.479[e]
		BEHAVE-AD - anxiety and phobias - 78wk	D	37	14(37.8%)	33	15(45.5%)	31	15(48.4%)	0.687[e]	0.529[e]	0.988[e]

[a] MC=Mean Change; D=Dichotomous; TTE=time-to-event

[b] ANCOVA, covarying country, MMSE category, and baseline score

[c] Wilcoxon rank sum test

[d] "Wilcoxon's test"; unclear whether method adopted accounts for right censorship of participants

[e] chi-square test (Yates's correction) (calculated by reviewer)

Synthesis with existing evidence-base

Once more, heterogeneity of measures reported and follow-up times at which data are available makes it impossible to perform meaningful synthesis within or between the newly identified evidence base and that reported in 2004.

4.7.1.2.4. Global effect

New data

Bullock and colleagues used the GDS to measure overall effect.[142] They found that, over the two-year trial, individuals who had been randomised to donepezil deteriorated by around 0.1 points more than those taking rivastigmine. As with the difference found on their chosen functional measure, this discrepancy appeared significant in the ITT LOCF analysis ($p<0.05$ by Wilcoxon rank sum test), but this finding was not repeated in secondary analyses based on evaluable and observed cases.

In the Ancoli-Israel and colleagues' RCT,[143] none of the individuals taking galantamine experienced a global decline, according to the CIBIC plus, over the eight weeks of treatment, whereas 13% of those taking donepezil deteriorated on the same measure, although this difference does not appear to be a significant one (*Table 38*).

TABLE 39 *Measures of global effect in included studies – head-to-head comparisons*

Study	Subgroup	Outcome	Type	Rivastigmine		Donepezil		Galantamine		*p*	
				N	Mean / n (%)	N	Mean / n (%)	N	Mean / n (%)	R v. D	D v. G
Bullock et al. (2005)[142]	LOCF analysis	GDS - 104wk	mean change	471	0.58 (SD 0.9)	483	0.69 (SD 0.9)	-	-	0.049[a]	-
Ancoli-Israel et al. (2005)[143]	OC population	CIBIC-plus score - 8wk	absolute value	-	-	29	3.97 (SD 1.02)	27	3.59 (SD 0.64)	-	0.106[b]
		CIBIC-plus: markedly improved - 8wk	dichotomous	-	-	29	0 (0.0%)	27	0 (0.0%)	-	0.330[c]
		CIBIC-plus: moderately improved - 8wk	dichotomous	-	-	29	3 (10.3%)	27	2 (7.4%)	-	0.933[c]
		CIBIC-plus: minimally improved - 8wk	dichotomous	-	-	29	4 (13.8%)	27	7 (25.9%)	-	0.421[c]
		CIBIC-plus: no change - 8wk	dichotomous	-	-	29	18 (62.1%)	27	18 (66.7%)	-	0.936[c]
		CIBIC-plus: minimally worse - 8wk	dichotomous	-	-	29	3 (10.3%)	27	0 (0.0%)	-	0.334[c]
		CIBIC-plus: moderately worse - 8wk	dichotomous	-	-	29	3 (10.3%)	27	0 (0.0%)	-	0.334[c]
		CIBIC-plus: markedly worse - 8wk	dichotomous	-	-	29	0 (0.0%)	27	0 (0.0%)	-	0.330[c]

[a] Wilcoxon rank sum test

[b] student's t-test (calculated by reviewer)

[c] chi-square test (Yates's correction) (calculated by reviewer)

Synthesis with existing evidence-base

Quantitative synthesis combining newly identified evidence and/or that reported in 2004 was not possible, due to heterogeneity of measures reported and follow-up times at which data are available.

4.7.1.2.5. Quality of life

None of the newly identified, head-to-head, randomised studies investigated QoL with the technologies under assessment, and no such data were identified in the 2004 review.

4.7.1.2.6. Safety

A variety of AEs were reported in the included studies, the most common were nausea, diarrhoea, vomiting and headache (*Table 40*, *Table 41* and *Table 42).*

TABLE 40 *AEs in included head-to-head studies – donepezil* v. *galantamine*

Adverse event	Donepezil						Galantamine						DvG		
	Ancoli-Israel et al. (2005)[143]		Cumbo (2005)[92]		Nordberg et al. (2009)[141]		Ancoli-Israel et al. (2005)[143]		Cumbo (2005)[92]		Nordberg et al. (2009)[141]		Ancoli-Israel et al. (2005)[143]	Cumbo (2005)[92]	Nordberg et al. (2009)[141]
	N	n (%)	N	n (%)	N	n (%)	N	n (%)	N	n (%)	N	n (%)	*p*[a]		
Abdominal pain					20	2 (10.0%)					21	0 (0.0%)			0.447
Anorexia			31	0 (0.0%)					33	1 (3.0%)				0.975	
Bronchitis	32	0 (0.0%)					31	3 (9.7%)					0.226		
Constipation	32	3 (9.4%)					31	0 (0.0%)					0.248		
Diarrhoea	32	5 (15.6%)			20	0 (0.0%)	31	1 (3.2%)			21	6 (28.6%)	0.212		0.032
Dizziness					20	1 (5.0%)					21	3 (14.3%)			0.635
Headache	32	3 (9.4%)	31	2 (6.5%)	20	2 (10.0%)	31	2 (6.5%)	33	0 (0.0%)	21	2 (9.5%)	0.970	0.445	0.635
Influenza					20	0 (0.0%)					21	2 (9.5%)			0.490
Injury	32	2 (6.3%)					31	2 (6.5%)					0.628		
Insomnia					20	2 (10.0%)					21	2 (9.5%)			0.635
Muscle spasms					20	3 (15.0%)					21	1 (4.8%)			0.563
Nausea	32	1 (3.1%)	31	2 (6.5%)	20	2 (10.0%)	31	3 (9.7%)	33	2 (6.1%)	21	6 (28.6%)	0.583	0.651	0.269
Pain[b]	32	3 (9.4%)					31	2 (6.5%)					0.970		
URTI					20	1 (5.0%)					21	0 (0.0%)			0.980
Vomiting			31	0 (0.0%)	20	0 (0.0%)			33	1 (3.0%)	21	3 (14.3%)		0.975	0.248
Weight decrease			31	0 (0.0%)					33	1 (3.0%)				0.975	
Weight loss					20	1 (5.0%)					21	1 (4.8%)			0.490

[a] chi-squared test (Yates's correction), calculated by reviewer

[b] no further detail provided

TABLE 41 *AEs in included head-to-head studies – donepezil* v. *rivastigmine*

Adverse event	Donepezil						Rivastigmine						DvR		
	Bullock et al. (2005)[142]		Cumbo (2005)[92]		Nordberg et al. (2009)[141]		Bullock et al. (2005)[142]		Cumbo (2005)[92]		Nordberg et al. (2009)[141]		Bullock et al. (2005)[142]	Cumbo (2005)[92]	Nordberg et al. (2009)[141]
	N	n (%)	N	n (%)	N	n (%)	N	n (%)	N	n (%)	N	n (%)	p[a]		
Any AE	453	349 (77.0%)					404	318 (78.7%)					0.613		
Any serious AE	499	162 (32.5%)					495	157 (31.7%)					0.854		
Abdominal pain					20	2 (10.0%)					22	0 (0.0%)			0.427
Aggression	453	25 (5.5%)					404	19 (4.7%)					0.700		
Agitation	453	47 (10.4%)					404	34 (8.4%)					0.389		
Anorexia	453	14 (3.1%)	31	0 (0.0%)			404	26 (6.4%)	37	1 (2.7%)			**0.031**	0.929	
Depression	453	16 (3.5%)					404	21 (5.2%)					0.303		
Diarrhoea	453	30 (6.6%)			20	0 (0.0%)	404	26 (6.4%)			22	2 (9.1%)	0.978		0.512
Dizziness					20	1 (5.0%)					22	3 (13.6%)			0.67
Fall	453	44 (9.7%)					404	33 (8.2%)					0.503		
Headache	453	12 (2.6%)	31	2 (6.5%)	20	2 (10.0%)	404	13 (3.2%)	37	1 (2.7%)	22	3 (13.6%)	0.771	0.875	0.91
Hypertension	453	18 (4.0%)					404	21 (5.2%)					0.487		
Influenza					20	0 (0.0%)					22	1 (4.5%)			0.962
Injury															
Insomnia					20	2 (10.0%)					22	1 (4.5%)			0.932
Muscle spasms					20	3 (15.0%)					22	0 (0.0%)			0.199
Nausea	453	24 (5.3%)	31	2 (6.5%)	20	2 (10.0%)	404	52 (12.9%)	37	3 (8.1%)	22	10 (45.5%)	**<0.001**	0.837	0.028
Upper respiratory tract infection					20	1 (5.0%)					22	2 (9.1%)			0.932
Urinary tract infection	453	26 (5.7%)					404	18 (4.5%)					0.487		
Vomiting	453	20 (4.4%)	31	0 (0.0%)	20	0 (0.0%)	404	62 (15.3%)	37	1 (2.7%)	22	4 (18.2%)	**<0.001**	0.929	0.139
Weight decrease	453	43 (9.5%)	31	0 (0.0%)			404	36 (8.9%)	37	0 (0.0%)			0.861		
Weight loss					20	1 (5.0%)					22	2 (9.1%)			0.932

[a] chi-squared test (Yates's correction), calculated by reviewer

TABLE 42 *AEs in included head-to-head studies – galantamine* v. *rivastigmine*

Adverse event	Galantamine				Rivastigmine				GvR	
	Cumbo (2005)[92]		Nordberg et al. (2009)[141]		Cumbo (2005)[92]		Nordberg et al. (2009)[141]		Cumbo (2005)[92]	Nordberg et al. (2009)[141]
	N	n (%)	N	n (%)	N	n (%)	N	n (%)	*p*[a]	
Abdominal pain			21	0 (0.0%)			22	0 (0.0%)		-
Anorexia	33	1 (3.0%)			37	1 (2.7%)			0.524	
Diarrhoea			21	6 (28.6%)			22	2 (9.1%)		0.212
Dizziness			21	3 (14.3%)			22	3 (13.6%)		0.705
Headache	33	0 (0.0%)	21	2 (9.5%)	37	1 (2.7%)	22	3 (13.6%)	0.954	0.956
Influenza			21	2 (9.5%)			22	1 (4.5%)		0.967
Insomnia			21	2 (9.5%)			22	1 (4.5%)		0.967
Muscle spasms			21	1 (4.8%)			22	0 (0.0%)		0.981
Nausea	33	2 (6.1%)	21	6 (28.6%)	37	3 (8.1%)	22	10 (45.5%)	0.894	0.407
Upper respiratory tract infection			21	0 (0.0%)			22	2 (9.1%)		0.490
Vomiting	33	1 (3.0%)	21	3 (14.3%)	37	1 (2.7%)	22	4 (18.2%)	0.524	0.946
Weight decrease	33	1 (3.0%)			37	0 (0.0%)			0.954	
Weight loss			21	1 (4.8%)			22	2 (9.1%)		0.967

[a] chi-squared test (Yates's correction), calculated by reviewer

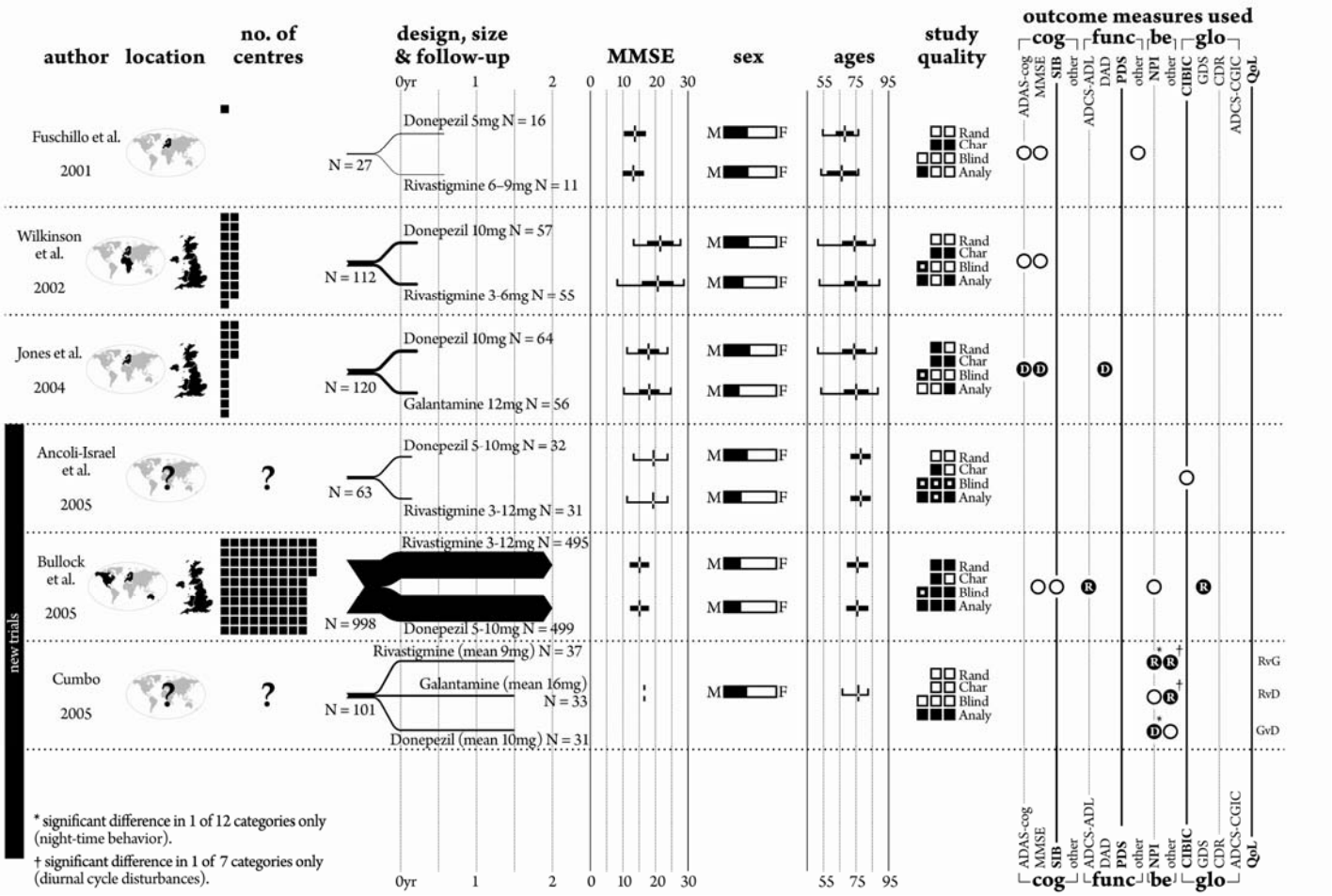
4.7.1.3. Summary: head-to-head comparisons

Four new head to head RCTs were found; two compared all included AChEIs, one compared donepezil to rivastigmine and one compared donepezil to galantamine. Pooling of data from head-to-head trials was not possible due to the heterogeneity of the data. The quality of the evidence they provide is limited due to the poor quality of most of the trials. The exception to this was Bullock and colleagues whose good quality study found no significant difference between donepezil and rivastigmine for cognitive or behavioural outcomes. However, when they looked at functional and global outcomes, patients taking rivastigmine faired significantly better than those taking donepezil in the primary analysis.

4.7.1.4. Graphical summary of head-to-head comparisons

Figure 51 clearly shows that this group of studies is dominated by the new comparison of rivastigmine with donepezil by Bullock and colleagues.[142] The small studies from the previous review indicate that there is no difference between donepezil and galantamine on cognitive outcomes. These earlier results support those of the much larger study by Bullock and colleagues, which also shows that rivastigmine is significantly better than donepezil on functional and global outcomes. Previously, when donepezil and galantamine were compared in a small poor quality trial the results favoured donepezil on cognitive and functional outcomes; no new evidence for this comparison for these outcomes was found. There was no good or even moderate evidence comparing all three AChEIs.

FIGURE 51 *Summary of all included head-to-head studies in 2004 and 2010*



4.8. Combination therapy

4.8.1.1. Identified evidence

Our searches identified a single new randomised trial addressing the effectiveness of combination therapy consisting of two of the technologies under review. Details of design and characteristics are presented in *Table 43* and *Table 44* and an assessment of study quality in *Table 45.*

One included study in the 2004 review – Tariot and colleagues. (2004)[135] – addressed the effectiveness of donepezil+memantine *v.* donepezil+placebo. In the 2004 review, this is considered among the evidence of effectiveness of memantine. We have not followed this approach, as we prefer to assess monotherapy and combination regimens separately, because the effect of multiple agents may or may not be straightforwardly additive.

TABLE 43 *Design of included studies – combination therapy*

Study details	Inclusion criteria	Exclusion criteria	Methodological notes	Other
Porsteinsson et al. (2008)[144] **Design:** Parallel double-blind RCT **Country:** USA **No. of centres:** 38 **No. randomised:** 433 **Maximum follow-up:** 24 **MMSE range included:** 10–22	Probable AD (NINCDS-ADRDA criteria) Age >=50y MRI or CT scan results consistent with AD diagnosis and acquired within 1y of study Treatment with cholinesterase inhibitors for >=6mo, and a stable dosing regimen for >=3mo (donezepil 5 or 10mg/ day; rivastigmine 6, 9 or 12 mg/day; galantamine 16 or 24mg/day) A knowledgable and reliable caregiver to acompany the participant to all study visits and supervise administraton of study frug Ability to ambulate Vision and hearing sufficient to permit compliance with	Clinically significant and active pulmonary, gastrointestinal, renal, hepatic, endocrine, or cardiovascular disease Clinically significant B12 or folate deficiency Evidence (including CT/MRI) of other psychiatric or neurological disorders Dementia complicated by organic disease or AD with delusions or delirium Undergoing treatment for an oncology diagnosis, or completion of treatment within 6mo of screening Modified Hachinski Ischaemia Scale score >4 Poorly controlled hypertension Substance abuse Participation in an investigational drug	**Sample attrition / dropout:** 385 of 433 completed study. Drop-outs in memantine arm: adverse events n=13, withdrew consent n=4, protocol violation n=5, insufficient therapeutic response n=1; drop-outs in placebo arm: adverse events n=17, withdrew consent n=4, protocol violation n=1, insufficient therapeutic response n=1, other n=2. No differences between groups. **Randomisation and allocation:** Randomised in permuted blocks of 4 in accordance with randomisation list generated and retained by Forest Research Institute, Department of Statistical Programming. Participants were sequentially assigned randomisation numbers at the baseline visit. No individual participant randomisation code was revealed during the trial. Memantine and placebo tablets described as being identical in appearance. **Power calculation:** Assuming an effect size (defined as difference of mean scores between treatment groups on ADAS-Cog at endpoint (LOCF), relative to pooled standard deviation) of 0.325, at least 400 participants	**Therapy common to all participants:** all participants continued to take cholinesterase inhibitor (donepezil, galantamine, or rivastigmine) 1 to 2 week single-blind placebo lead-in phase completed before randomisation to assess compliance **Study Funding:** Forest Laboratories, Inc. (New York, NY) provided all financial and material support for research and analyses – and assisted the Memantine Study Group in the development of the trial design, implementation, data collection, post-hoc analyses, and manuscript development. **Other conflicts:** One co-author's (JO) affilliation is Novartis, Inc.

Study details	Inclusion criteria	Exclusion criteria	Methodological notes	Other
	assessments Montgomery-Asberg Depression Rating Scale (MADRS) score <22 Medically stable Post-menopausal for >=2yr, or surgically sterile (female participants)	study or use of an investigational drug within 30dy (or 5 half-lives, whichever is longer) of screening Depot neuroleptic use within 6mo of screening Positive urine drug test Likely institutionalisation during trial Previous memantine treatment or participation in an investgational study of memantine Likely cessation of cholinesterase inhibitors during the trial	were needed to provide 90% power at an alpha level of 0.05 (2-sided), based on a 2-sided t test. The total patient population, consisting of all participants randomised into the study (n=433) was identical to the safety population , which consusted of randomised participants who received at least 1 dose of double-blind study medication. The ITT population (n=427) comprised participants in the safety population who completed at least 1 post-baseline ADAS-cog or CIBIC-Plus assessment.	

TABLE 44 *Interventions, comparators, and baseline characteristics of participants in included studies – memantine & cholinesterase inhibitors* v. *cholinesterase inhibitors*

Study	Arm	Dose (mg/d)	Dosage details	N	Age	Sex (n male)	Race (n white)	Weight (kg)	Education (yrs)	Duration of dementia mo)	ADAS-cog	MMSE
Porsteinsson et al. (2008)[144]	Memantine	5–20	Titrated from an initial dosage of 5mg/d in 5mg weekly increments to a maximum dose of 20mg/d (administered as four 5mg tablets qd at bedtime) AChEIs administered separately as part of participants' ongoing maintenance therapy.	217	74.9 (SD 7.64)	100 (46.1%)		70.0 (SD 14.9)				16.7 (SD 3.67)
	Placebo	-	Placebo had identical appearance to memantine. AChEIs administered separately as part of participants' ongoing maintenance therapy.	216	76.0 (SD 8.43)	107 (49.5%)		72.2 (SD 14.7)				17.0 (SD 3.64)

TABLE 45 *Markers of internal validity of included studies – memantine & cholinesterase inhibitors* v. *cholinesterase inhibitors*

	Was the assignment to the treatment groups really random?	Was the treatment allocation concealed?	Were the groups similar at baseline in terms of prognostic factors?	Were the eligibility criteria specified?	Were outcome assessors blinded to the treatment allocation?	Was the care provider blinded?	Was the patient blinded?	Were the point estimates and measure of variability presented for the primary outcome measure?	Did the analyses include an intention-to-treat analysis?	Were withdrawals and dropouts completely described?
Porsteinsson *et al.* (2008)[144]	ADEQUATE	ADEQUATE	REPORTED – YES	INADEQUATE	UNKNOWN	ADEQUATE	ADEQUATE	ADEQUATE	ADEQUATE	ADEQUATE

4.8.1.2. Evidence of clinical effectiveness

4.8.1.2.1. Cognition

The new study by Porsteinsson and colleagues (2008)[144] failed to show any benefit on cognitive outcomes from combining memantine with an AChEI (*Table 46)*. One reason for this may be due to an underlying pharmacological interaction between galantamine and memantine that could neutralise their effects.

TABLE 46 *Measures of cognition in included studies – combination therapy*

Study	Subgroup	Outcome	Type	AChEI + Memantine		AChEI + Placebo		*p*
				N	Mean	N	Mean	
Porsteinsson et al. (2008)[144]	LOCF analysis	ADAS-cog - 24wk	Continuous	214	28.5 (SD 12.8)	213	28 (SD 11.9)	0.184[a]
		MMSE - 24wk	Continuous	210	16.5 (SD 5.38)	198	16.4 (SD 5.08)	0.123[a]
	OC population	ADAS-cog - 24wk	Continuous	192	28.2 (SD 12.8)	188	27.6 (SD 11.7)	0.186[a]
		MMSE - 24wk	Continuous	193	16.6 (SD 5.41)	188	16.4 (SD 5.08)	0.190[a]

[a] ANCOVA (treatment group and centre as main effects; baseline score as covariate)

Synthesis with existing evidence-base

Because the previously identified study – Tariot et al. (2004)[135] – relies on SIB to estimate the effect of combination therapy on cognition, whereas Porsteinsson and colleagues[144] report MMSE and ADAS-cog, it was not possible to combine the two studies in a WMD meta-analysis, nor would it have been informative to combine two RCTs on a standardised scale.

4.8.1.2.2. Functional

New data

Similarly, functional outcomes failed to show a significant difference between combination therapy and an AChEI plus placebo (*Table 47)*.

TABLE 47 *Measures of functional ability in included studies – combination therapy*

Study	Subgroup	Outcome	Type	AChEI + Memantine		AChEI + Placebo		*p*
				N	Mean	N	Mean	
Porsteinsson et al. (2008)[144]	LOCF analysis	ADCS-ADL - 24wk	Continuous	214	51.8 (SD 15.9)	213	52 (SD 15.7)	0.816[a]
	OC population	ADCS-ADL - 24wk	Continuous	193	51.8 (SD 16)	189	53.6 (SD 14.6)	0.741[a]

[a] ANCOVA (treatment group and centre as main effects; baseline score as covariate)

Synthesis with existing evidence-base

Although both relevant studies use the ADCS-ADL to measure the functional effectiveness of combination therapy, different versions of the instrument are adopted: Porsteinsson and colleagues[144] rely on the 23-item scale, whereas Tariot and colleagues use the 19-item version. Accordingly, it is not valid to synthesise these data on their original scales.

4.8.1.2.3. Behavioural and mood

New data

Porteinsson and colleagues also failed to show any benefit from combination therapy when behavioural outcomes were measured with the NPI (*Table 48).*

TABLE 48 *Measures of behavioural effect and mood in included studies – combination therapy*

Study	Subgroup	Outcome	Type	AChEI + Memantine		AChEI + Placebo		*p*
				N	Mean	N	Mean	
Porsteinsson et al. (2008)[144]	LOCF analysis	NPI - 24wk	Mean change	212	0.70 (SD 12.01)[a]	209	0.40 (SD 12.29)[a]	NS[b]
			Continuous	212	12.9 (SD 14.5)	209	12.6 (SD 14.6)	0.743[b]
	OC population	NPI - 12wk	Mean change	193	0.80 (SD 10.77)[a]	189	0.30 (SD 10.65)[a]	NS[b]
		NPI - 24wk	Mean change	193	0.00 (SD 11.81)[a]	189	0.00 (SD 11.69)[a]	NS[b]
		NPI - 24wk	Continuous	193	12.3 (SD 13.7)	189	11.9 (SD 13.5)	0.985[b]

[a] data estimated from figure

[b] ANCOVA (treatment group and centre as main effects; baseline score as covariate)

Synthesis with existing evidence-base

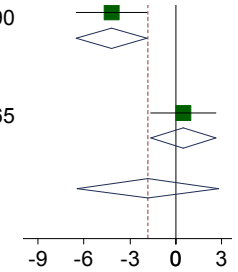
NPI

When the evidence from Porsteinsson and colleagues was pooled with the existing evidence from the NPI at 12 and 24 weeks, the overall pooled estimates showed no significant gain from combination therapy (*Figure 52* and *Figure 53).*

FIGURE 52 *Random-effects meta-analysis: NPI at 12wk (mean change from baseline) – AChEI+memantine v. AChEI+placebo*

	AChEI + Memantine			AChEI + Placebo				
	N	mean	SD	N	mean	SD	WMD (95%CI)	Wght
LOCF analysis								
Tariot et al. (2004)[135]	193	-2.50	11.00	186	1.70	11.90	-4.200 (-6.509, -1.891)	49.6
subtotal							**-4.200 (-6.509, -1.891)**	**49.6**
							p<0.001	
OC population								
Porsteinsson et al. (2008)[144]	193	0.80	10.77	189	0.30	10.65	0.500 (-1.648, 2.648)	50.4
subtotal							**0.500 (-1.648, 2.648)**	**50.4**
							p=0.648	
Overall pooled estimate							**-1.830 (-6.436, 2.776)**	
							p=0.436	

(Q=8.53 [p on 1 d.f.=0.003]; I^2=88.3%; τ^2=9.750)
Inter-stratum heterogeneity: p=0.003
Small-study effects: not calculable

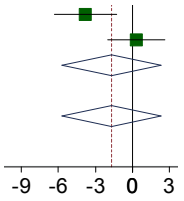


favours AChEI + memantine *favours AChEI + placebo*

FIGURE 53 *Random-effects meta-analysis: NPI at 24wk (mean change from baseline) – AChEI+memantine v. AChEI+placebo*

	AChEI + Memantine			AChEI + Placebo				
	N	mean	SD	N	mean	SD	WMD (95%CI)	Wght
LOCF analysis								
Tariot et al. (2004)[135]	193	-0.10	11.20	189	3.70	14.00	-3.800 (-6.346, -1.254)	49.2
Porsteinsson et al. (2008)[144]	212	0.70	12.01	209	0.40	12.29	0.300 (-2.022, 2.622)	50.8
subtotal (Q=5.44 [p on 1 d.f.=0.020]; I^2=81.6%; τ^2=6.860)							**-1.715 (-5.733, 2.302)**	**100.0**
							p=0.403	
Overall pooled estimate							**-1.715 (-5.733, 2.302)**	
							p=0.403	

(Q=5.44 [p on 1 d.f.=0.020]; I^2=81.6%; τ^2=6.860)
Small-study effects: not calculable



favours AChEI + memantine *favours AChEI + placebo*

4.8.1.2.4. Global effect

New data

Again, with global outcomes, no additional benefit was found from combination therapy (*Table 49).*

TABLE 49 *Measures of global effect in included studies – combination therapy*

Study	Subgroup	Outcome	Type	AChEI + Memantine		AChEI + Placebo		*p*
				N	Mean	N	Mean	
Porsteinsson et al. (2008)[144]	LOCF analysis	CIBIC-plus score - 24wk	Continuous	214	4.38 (SD 1)	213	4.42 (SD 0.96)	0.843[a]
	OC population	CIBIC-plus score - 24wk	Continuous	192	4.36 (SD 1.01)	189	4.4 (SD 0.96)	0.650[a]

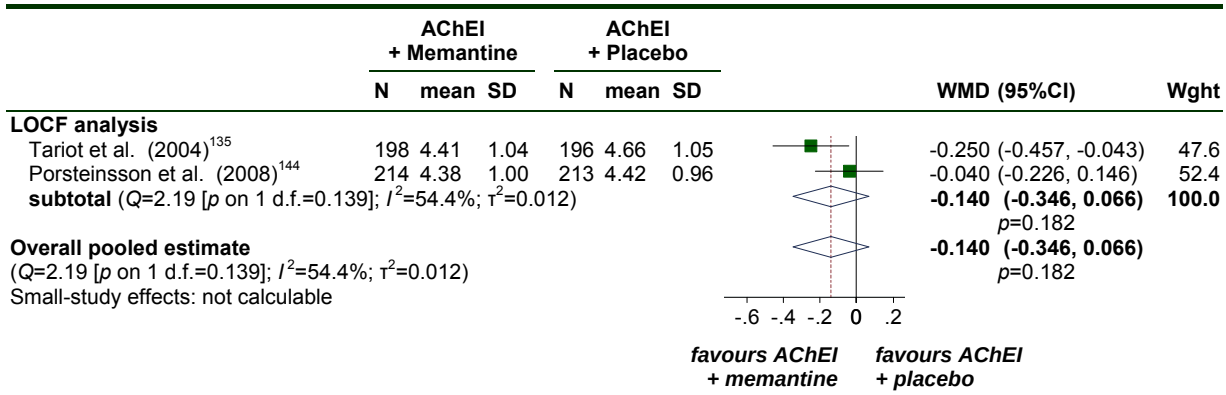
[a] Cochran–Mantel–Haenszel statistic using modified Ridit scores (Van Elteren test) controlling for study centre

Synthesis with existing evidence-base

Clinician Interview-Based Impression of Change- plus

A synthesis of new and existing evidence for global outcomes showed no overall benefit from combination therapy (*Table 46).*

FIGURE 54 *Random-effects meta-analysis: CIBIC-plus at 24wk – AChEI+memantine v. AChEI+placebo*



4.8.1.2.5. Quality of life

The new included study did not provide any randomised evidence on QoL with combination therapy, and no such data were identified in the 2004 review.

4.8.1.2.6. Safety

The proportion of AEs did not significantly vary between groups. The most common issues were falls and injury (*Table 50).*

TABLE 50 *AEs in included studies – combination therapy*

Adverse event	Porsteinsson et al. (2008)[144]				
	AChEI + Memantine		AChEI + Placebo		*p*
	N	n (%)	N	n (%)	
Any serious AE	217	27 (12.4%)	216	30 (13.9%)	0.762[a]
Diarrhoea	217	12 (5.5%)	216	14 (6.5%)	0.830[a]
Agitation	217	17 (7.8%)	216	17 (7.9%)	0.869[a]
Depression	217	14 (6.5%)	216	15 (6.9%)	0.990[a]
Injury	217	20 (9.2%)	216	16 (7.4%)	0.612[a]
Dizziness	217	16 (7.4%)	216	16 (7.4%)	0.865[a]
Upper respiratory tract infection	217	12 (5.5%)	216	6 (2.8%)	0.233[a]
Fall	217	22 (10.1%)	216	15 (6.9%)	0.309[a]
Influenza-like symptoms	217	15 (6.9%)	216	12 (5.6%)	0.700[a]
Abnormal gait	217	14 (6.5%)	216	9 (4.2%)	0.398[a]
Confusion	217	12 (5.5%)	216	9 (4.2%)	0.662[a]
Fatigue	217	11 (5.1%)	216	7 (3.2%)	0.476[a]
Hypertension	217	11 (5.1%)	216	6 (2.8%)	0.327[a]

[a] chi-square test (Yates's correction) (calculated by reviewer)

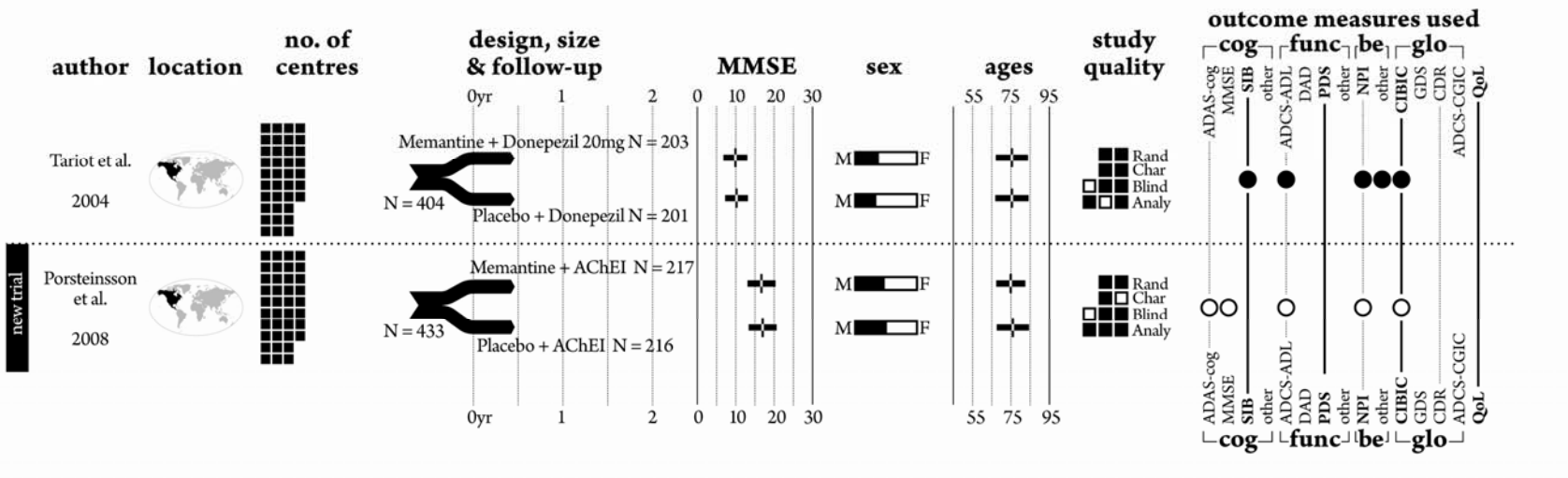
4.8.1.3. Summary: combination therapy

Our searches found one new trial that compared memantine plus an AChEI with an AChEI. This failed to show any benefit from combining memantine with an AChEI on cognitive, functional, behavioural or global outcomes. Pooling this data with previous trails also failed to show any additional benefit from combination therapy.

4.8.1.4. Graphical summary of combination therapy

The graphic below (*Figure 55*) clearly illustrates the similarities between the new and existing evidence for combination therapy; what is striking is the difference in results. Some of the variation may be explained by the use of different outcome measures or versions of outcome measures. However, it is unclear why the behavioural and global outcome results are different. The designs of these studies differed in that Porsteinsson and colleagues combined memantine with any of the three included AChEIs while Tariot and colleagues only combined memantine with donepezil. The other notable difference is that the 2004 authors analysed a modified ITT population whilst the 2008 study authors analysed a full ITT population. Whether these differences are sufficient to account for these differences in apparently similar populations is unknown.

FIGURE 55 *Summary of all studies included in 2004 and 2010 reviews – combination therapy*



4.9. Mixed treatment comparisons-indirect comparisons

Where there was sufficient data we pooled information on all technologies and their comparators simultaneously, in a MTC using Bayesian MCMC sampling.[83-86] The results are shown in terms of treatment effect compared to a common baseline. The evidence network shows the comparisons that were available and the quantity of those comparisons (by the thickness of the connecting lines). See Section *4.1.5.3* for more details. Mixed treatment comparisons of the technologies performed in specified measurement populations can be found in Appendix 9.

4.9.1. Cognitive

4.9.1.1. ADAS-cog

Table 51 shows the studies pooled in this MTC at 12–16 weeks follow-up, with their evidence network and effectiveness estimates. The results in *Table 52* give the relative effectiveness of each technology compared to placebo, indicating that donepezil and galantamine are certainly more effective than placebo and that donepezil is probably the most effective of these (0.48).

TABLE 51 *Mixed treatment comparison – ADAS-cog at 12–16wk (mean change from baseline; all measurement populations): input data*

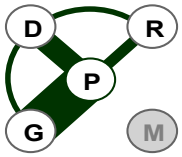
Evidence Network	Comparison	Study	WMD	(95%CI)
	Donepezil v. Placebo	Rogers et al. (1998)[107]	-2.799	(-3.831, -1.767)
		Burns et al. (1999)[97]	-2.151	(-2.871, -1.430)
		Homma et al. (2000)[101]	-2.175	(-3.527, -0.823)
		Nunez et al. (2003)[104;105]	-0.050	(-1.782, 1.682)
	Galantamine v. Placebo	Raskind et al. (2000)[117]	-3.158	(-4.371, -1.946)
		Tariot et al. (2000)[119]	-2.225	(-3.042, -1.408)
		Wilcock et al. (2000)[121]	-2.848	(-3.829, -1.867)
		Rockwood et al. (2001)[118]	-1.600	(-2.704, -0.496)
		Wilkinson & Murray (2001)[122]	-2.246	(-3.872, -0.620)
		Bullock et al. (2004)[94]	-1.475	(-2.933, -0.017)
		Brodaty et al. (2005)[89]	-2.453	(-3.192, -1.713)
		Rockwood et al. (2006)[125]	-1.925	(-3.816, -0.034)
	Rivastigmine v. Placebo	Feldman & Lane (2007)[132]	-2.249	(-3.226, -1.271)
		Winblad et al. (2007)[134]	-0.911	(-1.817, -0.006)
	Donepezil v. Rivastigmine	Wilkinson et al. (2002)[139]	0.150	(-1.561, 1.861)
	Donepezil v. Galantamine	Jones et al. (2004)[140]	-2.225	(-4.131, -0.319)

TABLE 52 *Mixed treatment comparison – ADAS-cog at 12–16wk (mean change from baseline; all measurement populations): results*

Technology	v. Placebo			Prob. most effective
	Effect	(95%CI)	Prob. more effective than placebo	
Placebo	-	-	-	0.000
Donepezil	-2.209	(-2.951, -1.452)	1.000	0.475
Galantamine	-2.176	(-2.725, -1.540)	1.000	0.421
Rivastigmine	-1.700	(-2.728, -0.751)	0.999	0.104
Memantine	-	-	-	-

At 21–26 weeks follow-up the MTC showed that all the treatments were more effective than placebo, with galantamine probably the most effective (0.89) (*Table 53* and *Table 54).*

TABLE 53 *Mixed treatment comparison – ADAS-cog at 21–26wk (mean change from baseline; all measurement populations): input data*

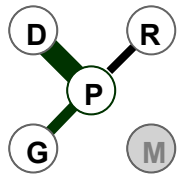
Evidence Network	Comparison	Study	WMD	(95%CI)
	Donepezil v. Placebo	Rogers et al. (1998)[106]	-2.684	(-3.876, -1.491)
		Burns et al. (1999)[97]	-2.203	(-2.968, -1.438)
		Homma et al. (2000)[101]	-2.540	(-3.427, -1.653)
	Galantamine v. Placebo	Raskind et al. (2000)[117]	-3.653	(-4.696, -2.611)
		Tariot et al. (2000)[119]	-2.741	(-3.633, -1.850)
		Wilcock et al. (2000)[121]	-3.049	(-4.030, -2.068)
		Bullock et al. (2004)[94]	-3.100	(-4.620, -1.580)
		Brodaty et al. (2005)[89]	-2.651	(-3.449, -1.854)
	Rivastigmine v. Placebo	Corey-Bloom et al. (1998)[129]	-2.751	(-3.694, -1.808)
		Rosler et al. (1999){Rosler, 1999 2016 /id}	-0.785	(-1.851, 0.281)
		Feldman & Lane (2007)[132]	-2.298	(-3.460, -1.137)
		Winblad et al. (2007)[134]	-1.943	(-2.858, -1.029)

TABLE 54 *Mixed treatment comparison – ADAS-cog at 21–26wk (mean change from baseline; all measurement populations): results*

Technology	v. Placebo			Prob. most effective
	Effect	(95%CI)	Prob. more effective than placebo	
Placebo	-	-	-	0.000
Donepezil	-2.431	(-3.174, -1.709)	1.000	0.107
Galantamine	-2.986	(-3.591, -2.405)	1.000	0.885
Rivastigmine	-1.978	(-2.630, -1.303)	1.000	0.009
Memantine	-	-	-	-

4.9.1.2. MMSE

The data used for the 12–13 week MTC for MMSE can be seen in *Table 55*. The results in *Table 56* show that at this early follow-up donepezil is the only treatment certainly more effective than placebo and consequently probably the most effective treatment overall (0.54).

TABLE 55 *Mixed treatment comparison – MMSE at 12–13wk (mean change from baseline; all measurement populations): input data*

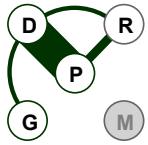
Evidence Network	Comparison	Study	WMD	(95%CI)
	Donepezil v. Placebo	Rogers et al. (1998)[107]	1.110	(0.514, 1.706)
		Mohs et al. (2001)[103]	1.600	(0.889, 2.311)
		Winblad et al. (2001)[110]	0.800	(0.075, 1.525)
		Gauthier et al. (2002)[98]	2.000	(0.820, 3.180)
		Nunez et al. (2003)[104;105]	0.830	(-0.071, 1.731)
		AD2000 (2004)[96]	0.930	(0.389, 1.471)
		Holmes et al. (2004)[100]	1.700	(0.169, 3.231)
		Seltzer et al. (2004)[109]	1.175	(0.100, 2.250)
	Rivastigmine v. Placebo	Agid et al. (1998)[128]	0.144	(-0.493, 0.782)
		Mowla et al. (2007)[133]	1.600	(1.099, 2.101)
	Donepezil v. Rivastigmine	Wilkinson et al. (2002)[139]	-0.490	(-1.825, 0.845)
	Donepezil v. Galantamine	Jones et al. (2004)[140]	0.888	(0.004, 1.771)

TABLE 56 *MTC – MMSE at 12–13wk (mean change from baseline; all measurement populations): results*

Technology	v. Placebo			Prob. most effective
	Effect	(95%CI)	Prob. more effective than placebo	
Placebo	-	-	-	0.000
Donepezil	1.145	(0.677, 1.637)	1.000	0.537
Galantamine	0.259	(-1.214, 1.761)	0.646	0.075
Rivastigmine	1.057	(0.283, 1.852)	0.993	0.389
Memantine	-	-	-	-

At 24–26 weeks from baseline there is no evidence from galantamine and donepezil continues to show that it is probably the most effective treatment (0.67).

TABLE 57 *MTC – MMSE at 24–26wk (mean change from baseline; all measurement populations): input data*

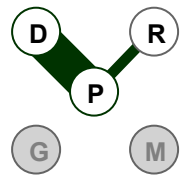
Evidence Network	Comparison	Study	WMD	(95%CI)
	Donepezil v. Placebo	Rogers et al. (1998)[106]	1.284	(0.604, 1.964)
		Mohs et al. (2001)[103]	1.350	(0.188, 2.512)
		Winblad et al. (2001)[110]	1.490	(0.548, 2.432)
		Gauthier et al. (2002)[98]	2.060	(0.880, 3.240)
		AD2000 (2004)[96]	0.500	(-0.250, 1.250)
		Seltzer et al. (2004)[109]	1.250	(0.171, 2.329)
		Mazza et al. (2006)[112]	1.450	(-3.720, 6.620)
	Rivastigmine v. Placebo	Feldman & Lane (2007)[132]	1.250	(0.670, 1.830)
		Winblad et al. (2007)[134]	0.932	(0.461, 1.403)

TABLE 58 *MTC – MMSE at 24–26wk (mean change from baseline; all measurement populations): results*

Technology	v. Placebo			Prob. most effective
	Effect	(95%CI)	Prob. more effective than placebo	
Placebo	-	-	-	0.000
Donepezil	1.235	(0.747, 1.778)	1.000	0.670
Galantamine	-	-	-	-
Rivastigmine	1.073	(0.358, 1.809)	0.993	0.330
Memantine	-	-	-	-

4.9.2. Functional

4.9.2.1. ADCS-ADL

Mixed treatment comparisons were conducted for the ADCS-ADL. At 12–16 weeks galantamine and rivastigmine were shown to be almost equally effective compared to placebo, with rivastigmine possibly being the most effective (0.50), (*Table 59* and *Table 60*).

TABLE 59 *MTC – ADCS-ADL at 12–16wk (mean change from baseline; all measurement populations): input data*

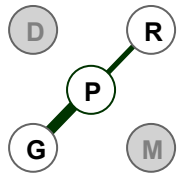
Evidence Network	Comparison	Study	WMD	(95%CI)
	Galantamine v. Placebo	Tariot et al. (2000)[119]	1.810	(0.613, 3.007)
		Brodaty et al. (2005)[89]	1.052	(-0.034, 2.138)
	Rivastigmine v. Placebo	Winblad et al. (2007)[134]	1.411	(0.279, 2.543)

TABLE 60 *MTC – ADCS-ADL at 12–16wk (mean change from baseline; all measurement populations): results*

Technology	v. Placebo			Prob. most effective
	Effect	(95%CI)	Prob. more effective than placebo	
Placebo	-	-	-	0.008
Donepezil	-	-	-	-
Galantamine	1.410	(-0.316, 3.148)	0.956	0.494
Rivastigmine	1.410	(-1.033, 3.842)	0.907	0.498
Memantine	-	-	-	-

At 21–26 weeks follow-up, the situation had changed with galantamine showing a slight greater probability of being the most effective technology (0.55), (*Table 61* and *Table 62).*

TABLE 61 *MTC – ADCS-ADL at 21–26wk (mean change from baseline; all measurement populations): input data*

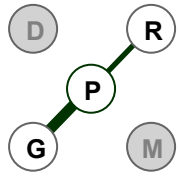
Evidence Network	Comparison	Study	WMD	(95%CI)
	Galantamine v. Placebo	Tariot et al. (2000)[119]	2.276	(0.889, 3.663)
		Brodaty et al. (2005)[89]	2.203	(1.007, 3.399)
	Rivastigmine v. Placebo	Winblad et al. (2007)[134]	2.101	(0.788, 3.415)

TABLE 62 *MTC – ADCS-ADL at 21–26wk (mean change from baseline; all measurement populations): results*

Technology	v. Placebo			Prob. most effective
	Effect	(95%CI)	Prob. more effective than placebo	
Placebo	-	-	-	0.001
Donepezil	-	-	-	-
Galantamine	2.238	(0.528, 3.943)	0.990	0.547
Rivastigmine	2.091	(-0.322, 4.519)	0.962	0.451
Memantine	-	-	-	-

4.9.3. Behavioural

4.9.3.1. NPI

The MTC for behavioural outcomes were measured using the NPI. At 12–13 weeks from baseline, donepezil was probably more effective than galantamine at controlling behavioural symptoms (0.78) (*Table 63* and *Table 64)*.

TABLE 63 *MTC – NPI at 12–13wk (mean change from baseline; all measurement populations): input data*

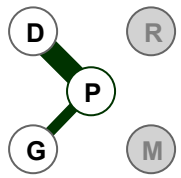
Evidence Network	Comparison	Study	WMD	(95%CI)
	Donepezil v. Placebo	Gauthier et al. (2002)[98]	-2.900	(-6.783, 0.983)
		Nunez et al. (2003)[104;105]	-2.870	(-5.406, -0.334)
		AD2000 (2004)[96]	1.250	(1.500, 4.000)
		Holmes et al. (2004)[100]	-6.200	(-11.374, -1.026)
	Galantamine v. Placebo	Tariot et al. (2000)[119]	-0.719	(-2.056, 0.618)
		Rockwood et al. (2001)[118]	-0.900	(-2.688, 0.888)

TABLE 64 *MTC – NPI at 12–13wk (mean change from baseline; all measurement populations): results*

Technology	v. Placebo			Prob. most effective
	Effect	(95%CI)	Prob. more effective than placebo	
Placebo	-	-	-	0.006
Donepezil	-1.960	(-4.095, 0.033)	0.973	0.799
Galantamine	-0.788	(-2.872, 1.267)	0.810	0.195
Rivastigmine	-	-	-	-
Memantine	-	-	-	-

At 21–28 weeks follow-up there was also data on rivastigmine and memantine to put into the MTC. However, donepezil was still probably the most effective treatment (0.57) (*Table 65* and *Table 66).*

TABLE 65 *MTC – NPI at 21–28wk (mean change from baseline; all measurement populations [all are classical ITT or LOCF analysis]): input data*

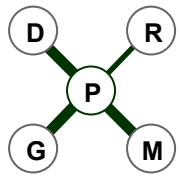
Evidence Network	Comparison	Study	WMD	(95%CI)
	Donepezil v. Placebo	Gauthier et al. (2002)[98]	-5.920	(-10.126, -1.714)
		AD2000 (2004)[96]	-0.750	(-3.750, 2.250)
	Galantamine v. Placebo	Tariot et al. (2000)[119]	-1.574	(-3.226, 0.078)
		Brodaty et al. (2005)[89]	-1.349	(-2.900, 0.202)
	Rivastigmine v. Placebo	Winblad et al. (2007)[134]	-0.372	(-2.205, 1.461)
	Memantine v. Placebo	Reisberg et al. (2003)[136]	-3.300	(-7.334, 0.734)
		Van Dyck et al. (2007)[137]	-0.100	(-3.845, 3.645)

TABLE 66 *MTC – NPI at 21–28wk (mean change from baseline; all measurement populations [all are classical ITT or LOCF analysis]): results*

Technology	v. Placebo			Prob. most effective
	Effect	(95%CI)	Prob. more effective than placebo	
Placebo	-	-	-	0.000
Donepezil	-2.683	(-5.673, 0.207)	0.966	0.576
Galantamine	-1.462	(-3.438, 0.526)	0.940	0.129
Rivastigmine	-0.366	(-3.308, 2.554)	0.612	0.052
Memantine	-1.600	(-4.762, 1.540)	0.845	0.243

4.9.4. Global

For global outcomes MTC was carried out using the CIBIC-plus and the GDS.

4.9.4.1. CIBIC-plus

At 12–16 weeks post-baseline a MTC of all the treatments showed that galantamine was probably the most effective treatment (0.54) (*Table 67* and *Table 68).*

TABLE 67 *Mixed treatment comparison – CIBIC-plus at 12–16wk (all measurement populations): input data*

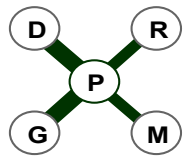
Evidence Network	Comparison	Study	WMD	(95%CI)
	Donepezil v. Placebo	Rogers et al. (1998)[107]	-0.350	(-0.527, -0.174)
		Burns et al. (1999)[97]	-0.265	(-0.406, -0.125)
		Gauthier et al. (2002)[98]	-0.490	(-0.768, -0.212)
	Galantamine v. Placebo	Rockwood et al. (2001)[118]	-0.335	(-0.524, -0.146)
		Rockwood et al. (2006)[125]	-0.450	(-0.797, -0.103)
	Rivastigmine v. Placebo	Rosler et al. (1999){Rosler, 1999 2016 /id}	-0.007	(-0.186, 0.172)
	Memantine v. Placebo	Reisberg et al. (2003)[136]	-0.070	(-0.347, 0.207)

TABLE 68 *MTC – CIBIC-plus at 12–16wk (all measurement populations): results*

Technology	v. Placebo			Prob. most effective
	Effect	(95%CI)	Prob. more effective than placebo	
Placebo	-	-	-	0.001
Donepezil	-0.338	(-0.647, -0.079)	0.985	0.373
Galantamine	-0.370	(-0.746, -0.025)	0.978	0.541
Rivastigmine	-0.007	(-0.492, 0.477)	0.520	0.027
Memantine	-0.071	(-0.591, 0.448)	0.647	0.058

However, at 24-28 weeks follow-up donepezil was probably the most effective (0.55) (*Table 69* and *Table 70).*

TABLE 69 *MTC – CIBIC-plus at 24–28wk (all measurement populations): input data*

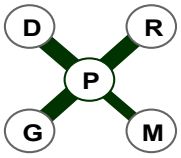
Evidence Network	Comparison	Study	WMD	(95%CI)
	Donepezil v. Placebo	Rogers et al. (1998)[106]	-0.400	(-0.593, -0.207)
		Burns et al. (1999)[97]	-0.340	(-0.484, -0.196)
		Gauthier et al. (2002)[98]	-0.545	(-0.858, -0.232)
	Galantamine v. Placebo	Raskind et al. (2000)[117]	-0.248	(-0.419, -0.077)
		Wilcock et al. (2000)[121]	-0.288	(-0.450, -0.127)
		Brodaty et al. (2005)[89]	-0.138	(-0.294, 0.018)
	Rivastigmine v. Placebo	Corey-Bloom et al. (1998)[129]	-0.275	(-0.471, -0.079)
		Rosler et al. (1999){Rosler, 1999 2016 /id}	-0.300	(-0.519, -0.081)
		Feldman & Lane (2007)[132]	-0.500	(-0.711, -0.289)
	Memantine v. Placebo	Reisberg et al. (2003)[136]	-0.300	(-0.582, -0.018)
		Van Dyck et al. (2007)[137]	-0.300	(-0.515, -0.085)

TABLE 70 *Mixed treatment comparison – CIBIC-plus at 24–28wk (all measurement populations): results*

Technology	v. Placebo			Prob. most effective
	Effect	(95%CI)	Prob. more effective than placebo	
Placebo	-	-	-	0.000
Donepezil	-0.392	(-0.549, -0.251)	1.000	0.546
Galantamine	-0.222	(-0.356, -0.091)	0.997	0.010
Rivastigmine	-0.354	(-0.508, -0.203)	1.000	0.285
Memantine	-0.300	(-0.507, -0.100)	0.996	0.159

4.9.4.2. Global Deterioration Scale

There was only data from the GDS at 24–28 weeks from baseline. This indicated that rivastigmine was probably more effective than donepezil or memantine for global outcomes (0.49), (*Table 71* and *Table 72).*

TABLE 71 *MTC – GDS at 24–28wk (mean change from baseline; all measurement populations): input data*

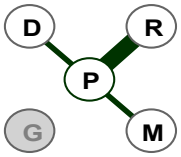
Evidence Network	Comparison	Study	WMD	(95%CI)
	Donepezil v. Placebo	Winblad et al. (2001)[110]	0.160	(-0.006, 0.326)
	Rivastigmine v. Placebo	Corey-Bloom et al. (1998)[129]	0.175	(0.065, 0.285)
		Rosler et al. (1999){Rosler, 1999 2016 /id}	0.120	(-0.042, 0.282)
		Feldman & Lane (2007)[132]	0.200	(0.087, 0.312)
	Memantine v. Placebo	Reisberg et al. (2003)[136]	-0.100	(-0.220, 0.020)

TABLE 72 *MTC – GDS at 24–28wk (mean change from baseline; all measurement populations): results*

Technology	v. Placebo			Prob. most effective
	Effect	(95%CI)	Prob. more effective than placebo	
Placebo	-	-	-	0.012
Donepezil	0.161	(-0.402, 0.720)	0.866	0.453
Galantamine	-	-	-	-
Rivastigmine	0.171	(-0.159, 0.486)	0.941	0.491
Memantine	-0.099	(-0.662, 0.450)	0.189	0.043

4.9.4.3. Summary: mixed treatment comparisons

The MTC results for cognitive outcomes varied with follow-up time and the measure used. Donepezil was shown to be probably the most effective treatment at short-term follow-up on the ADAS-cog and MMSE and this remained the case for the MMSE at 24 –26 weeks; however, the ADAS-cog favoured galantamine at this later follow-up time. Functional outcomes measured with the ADCS-ADL showed equal effectiveness from galantamine and rivastigmine at 12–16 weeks but by 21–26 weeks galantamine was probably the most effective treatment. For behavioural outcomes donepezil came out most favourably. For global outcomes the results were less clear, with galantamine probably the best treatment at 12–16 weeks when measured by the CIBIC-plus, but donepezil taking over by 24–28 weeks. However, when global outcomes were measured with the GDS rivastigmine came out as the most effective.

4.10. Summary of clinical effectiveness evidence

- From 1,843 titles and abstracts screened, four systematic reviews and 17 RCTs were found that matched our inclusion criteria that had been published since 2004.

- Overall the quality of the trials was disappointing, and there was insufficient evidence to suggest that one treatment is better than another. We therefore suggest that the AChEis are taken as a class of drugs.

- When combined with data from the previous review in 2004, donepezil was shown to provide gains on cognitive, functional and global outcomes when compared to placebo.

- Similar pooling of data from galantamine studies was conducted, showing clear benefits from cognitive, functional and global outcomes. Additionally, results favouring treatment were seen for behavioural outcomes at later (six month) follow-up.

- Pooled estimates of cognitive benefits from rivastigmine were favourable, but were shown to be dose–dependent as in the previous review. The results from functional and global outcomes also showed significant gains. However, results from individual trials of behavioural outcomes were mixed (pooling was not possible due to heterogeneity). The lower dose transdermal patch (9.5 mg/day) was shown to be as effective as the capsule (12 mg/day) but with fewer side-effects.

- The meta-analysis of memantine v. placebo showed benefit from memantine at 12 weeks follow-up on the SIB. However, treatment gain, measured by functional outcome, depended on the type of instrument used, and no benefit was seen from behavioural outcomes. Nevertheless, pooled estimates of global outcomes showed a benefit from taking memantine.

- Pooling of data from head-to-head trials was not possible due to the heterogeneity of the data. Results from the one reasonably good quality trial showed no significant difference between donepezil and rivastigmine for cognitive or behavioural outcomes. However, when looking at functional and global outcomes, patients taking rivastigmine fared significantly better than those taking donepezil in the primary analysis.

- Pooling data from trials combining memantine plus an AChEI v. an AChEI failed to show any additional benefit from combination therapy.

- The MTC results for cognitive outcomes varied. Donepezil and galantamine were both probably the most effective for cognitive outcomes depending on the measure used and the length of follow-up. Similarly, depending on follow-up time, galantamine and rivastigmine were either equally effective or galantamine was more effective on

functional outcomes. For longer term outcomes donepezil or rivastigmine were probably the most effective treatment depending on the measure used.

- As found in the previous review, the main adverse events for the AChEIs were gastrointestinal, and agitation and hypertension for memantine

The table below (*Table 73*), summarising the change in the evidence of effectiveness, as measured by statistical significance at p<0.05, should be interpreted with caution. Where there are fewer studies contributing to a category of outcome measure; e.g. cognitive, the more likely it is to have a positive or negative result and the more studies there are in a category the more likely it is that their results will go in different directions. Thus, there appears to be a possibly false sense of certainty about the memantine results when this may simply be an artefact of the number of studies.

TABLE 73 *Summary of the change in clinical effectiveness evidence since the 2004 review*

Outcome	Data	Donepezil (No. studies)	Galantamine (No. studies)	Rivastigmine (No. studies)	Memantine (No. studies)
Cognitive	New	~ (5)	✓ (3)	✓ (3)	✗ (1)
	Existing[145]	~ (6)	✓ (6)	~ (3)	~ (1)
	Pooled	✓	✓	✓	~[e]
Functional	New	✓ (1)	~ (3)	✓ (3)	✗(1)
	Existing	~ (8)	✓ (3)	~ (2)	✓ (1)
	Pooled	✓	✓	✓	~[f]
Behavioural	New	-	✗ (1)	✗ (2)	✗ (1)
	Existing	~ (4)	~ (2)	✗ (2)	✗ (1)
	Pooled	✗[g]	~[h]	-	✗
Global	New	✓ (1)	~ (2)	✓ (2)	✗ (1)
	Existing	~ (7)	✓ (5)	✓ (3)	✓ (1)
	Pooled	✓	✓	✓	✓
Change in direction of evidence	All	↑	↓	↑	↔
Change in amount of evidence	All	↑	↑	↑	↑
Increased precision	All	↑	↑	↑	↔[i]

~ The results of studies in this group were mixed for this outcome, some showing significant gain, others not

✓ The results of studies in this group all showed significant benefit (p<0.05) for this outcome

✗ The results of studies in this group all failed to show significant benefit (p<0.05) for this outcome

- This outcome was not measured for this drug

↑ Positive change in direction

↔ No change in direction

[e] The pooled results were significant at 12 weeks but not at 24-28 weeks follow-up.

[f] The pooled results were significant at 24-28 weeks with the FAST and the ADCS-ADL but not at 12 weeks with the ADCS-ADL.

[g] The pooled results were of existing studies

[h] The pooled results were significant at 21-26 weeks but not at 13 weeks follow-up.

[i] The quality of the new evidence was not as good as the previous evidence.

5. Assessment of cost-effectiveness

5.1. Introduction

The aim of this section is to assess the cost-effectiveness of donepezil, rivastigmine, galantamine and memantine for AD, updating the last guidance TA111 which considered evidence up to July 2004. The economic analysis comprises a systematic review of the literature on cost effectiveness, a review of the manufacturer's submissions to NICE on cost-effectiveness and this technology appraisal group's independent economic model.

The focus of the economic analysis is on evidence and analyses which have been produced since 2004. We do not review work which would already have been considered in previous technology assessments. Duplicate publications after 2004 of economic analyses or models originally published before 2004 (and included in the original economic analysis) would be clear examples of this.

5.1.1. Cost-effectiveness evidence which supported existing guidance

The starting point is thus the economic findings in the last guidance, which we summarise as follows. These have been taken from the text of NICE TA111 (amended September 2007, August 2009) Donepezil, galantamine, rivastigmine (review) and memantine for the treatment of AD (amended), recognising that this represents a summary of a highly complex series of deliberations over a number of years, starting with the TAR group report and industry submissions, incorporating additional analyses from NICE, considering responses from consultees, taking into account feed-back from a judicial review and then responses to this from the Decision Support Unit (DSU).

5.1.1.1. Concerning the acetylcholinesterase inhibitors (donepezil, galantamine and rivastigmine):

Main guidance:

> *"The three acetylcholinesterase inhibitors donepezil, galantamine and rivastigmine are recommended as options in the management of patients with Alzheimer's disease of moderate severity only (that is, subject to section 1.2 below, those with a Mini Mental State Examination [MMSE] score of between 10 and 20 points), and under specified conditions [as stated in 1.1]".*

Key economic considerations:

Different ICERs for moderate and mild AD were recognised. 4.3.10.8 states:

> *".....It (the appraisal committee) therefore considered whether it might be possible to define, prospectively, subgroups of people with Alzheimer's disease who might benefit more than average, and for whom AChE inhibitors might be a relatively cost-effective treatment ..... In accepting the subgroup analyses using severity of cognitive impairment, the Committee reviewed the estimates of cost effectiveness. It noted that for people with moderate Alzheimer's disease these estimates ranged from £23,000 to £35,000 depending on the choice of AChE inhibitor and by including carer benefits in the augmented base-case. Conversely, the Committee noted that for the subgroup of people with mild Alzheimer's disease estimates of cost effectiveness ranged from £56,000 to £72,000 depending on the choice of AChE inhibitor and by including carer benefits in the augmented base-case……."*

The specific ICERs given for moderate AD were:

> *"For moderate disease treated with donepezil, the augmented base-case ICER was £31,550 per QALY gained [from DSU augmented base-case as stated in 4.2.8.6]*
>
> *The impact of additional sensitivity analyses were also explored in the DSU analysis, but this did not appear to have a major effect on the ICERs for moderate AD [as stated in 4.2.8.8 to 4.2.11 inclusively]"*

Further detail on the ICERs for mild AD included :

> *"The Committee concluded that the cumulative impact of the changes it considered appropriate reduced the base-case ICER for mild Alzheimer's disease to approximately £55,000 to £58,000 per QALY gained (for galantamine and donepezil, respectively) which is further reduced by approximately £1500 when using the appropriate starting age of the full-time care index. The Committee noted the sensitivity analyses on estimates of health-related utility performed by the DSU but did not consider that the results of these were*

appropriate to consider as base-case estimates of the ICERs for the AChE inhibitors. It accepted that the ICERs could be lower than the base-case but concluded that the amendments had not reduced the ICERs for the subgroup of people with mild Alzheimer's disease to within the range normally accepted as a cost-effective use of NHS resources. [as stated in 4.3.37]"

5.1.1.2. Concerning memantine:

Main guidance:

"Memantine is not recommended as a treatment option for patients with moderately severe to severe Alzheimer's disease except as part of well-designed clinical studies [as stated in 1.4]"

Key economic considerations:

The main evidence on cost-effectiveness was derived from the model submitted by the manufacturer.

"In the probabilistic model submitted by the manufacturer, disease states were described by severity, level of dependency (dependent or independent), whether people were in institutional care or not and death. The people in the model made transitions between the states. The time horizon was 2 years. The transition probabilities between health states (defined as categories of MMSE score) were derived from a single RCT of memantine monotherapy. The odds ratio associated with institutionalisation was also derived from this single RCT and was not adjusted for differences in disease severity. The manufacturer calculated from this model that memantine dominated placebo for the total population as well as the subgroups except the subgroup of severe and dependent people with Alzheimer's disease for which an estimate of approximately £4000 was reported for the CQG. [as stated in 4.2.9.2]"

The influential estimates of cost-effectiveness were, however, generated by use of more plausible parameters in this model.

"The Assessment Group re-ran the model using a set of assumptions similar to those used in its own model for AChE inhibitors, and the CQG estimates were between £37,000 and £53,000. Further changes to transition probabilities in relation to the available trial evidence for, and costs of care associated with, memantine raised the estimated CQG in the manufacturer's model substantially above £53,000 [as stated in 4.2.9.3]"

This in turn led to the Committee concluding:

> *"….. on the basis of current evidence on clinical effectiveness memantine could not reasonably be considered a cost-effective therapy for moderately severe to severe Alzheimer's disease [as stated in 4.3.49]"*

The economic analysis for this report thus specifically considers whether new evidence would alter any of these conclusions and so lead the Appraisal Committee to reconsider the guidance. In this respect it should also be noted that the scope for memantine has changed slightly. In the previous guidance the licensed indication was in moderately-severe to severe AD, whereas for this report the licensed indication is moderate to severe AD.

5.2. Systematic review of existing economic evaluations

5.2.1. Method

General - This followed the process set out in the protocol from which there were no major departures.

This systematic review aimed to update the systematic review of cost-effectiveness studies which was conducted in 2004 as part of the review of evidence to inform the NICE's earlier guidance on these drugs (TA111).

The review aimed to summarise the main results of the included studies, and identify any key economic costs and trade-offs relevant to the decision problem. It also indicated the strengths and weaknesses of different modelling approaches in this treatment area.

It only fully extracted study data and assessed study quality for those economic evaluations or costing studies published since 2004 which were of relevance to the current decision problem. Further these were not to have duplicated work or analyses considered in the original guidance [last sentence added for clarification and did not appear in the original protocol].

Search strategy - The range of sources searched included those for clinical effectiveness and in addition NHSEED and Econlit. Full details of the search strategies are provided in Appendix 2.

Study selection criteria and procedures - The inclusion and exclusion criteria for the systematic review of economic evaluations were identical to those for the systematic review of clinical effectiveness, except:

- Non-randomised studies were included (e.g. decision model based analyses, or analyses of patient-level cost and effectiveness data alongside observational studies.)
- Full cost-effectiveness analyses, cost–utility analyses, cost-benefit analyses and cost consequence analyses were included. (Economic evaluations which only report average cost-effectiveness ratios will only be included if the incremental ratios can be easily calculated from the published data.)
- Standalone cost analyses based in the UK NHS were also sought and appraised.
- Based on the above inclusion/exclusion criteria, study selection were made by one reviewer (CH).
- Study quality assessment - The methodological quality of the economic evaluations were assessed according to internationally accepted criteria such as the City Health Economics Centre (CHEC) list questions developed by Evers and colleagues[146]. Any studies based on decision models were assessed against the International Society for Pharmacoeconomic and Outcomes Research (ISPOR) guidelines for good practice in decision analytic modelling[147].
- Data extraction strategy - For those studies which were of relevance to the current decision problem, data were extracted by one researcher (CH) into two summary tables: one to describe the study design of each economic evaluation and the other to describe the main results. [These have been merged for this report]

In the study design table the components were: author and year; model type or trial-based; study design (e.g. cost-effectiveness analysis [CEA], cost–utility analysis [CUA] or cost-analysis); service setting/country; study population; comparators; research question; perspective, time horizon, and discounting; main costs included; main outcomes included; sensitivity analyses conducted; and other notable design features.

For modelling-based economic evaluations a supplementary study design table recorded further descriptions of: model structure (and note its consistency with the study perspective,

and knowledge of disease/treatment processes); sources of transition & chance node probabilities; sources of utility values; sources of resource use and unit costs; handling of heterogeneity in populations; evidence of validation (e.g. debugging, calibration against external data, comparison with other models).

In the results table the components were: for each comparator, incremental cost; incremental effectiveness/utility and ICERs. Excluded comparators on the basis of dominance or extended dominance will also be noted. The original authors' conclusions were noted, and also any issues raised concerning the generalisability of results. Finally the reviewers' comments on study quality and generalisability (in relation to the TAR scope) of their results were recorded.

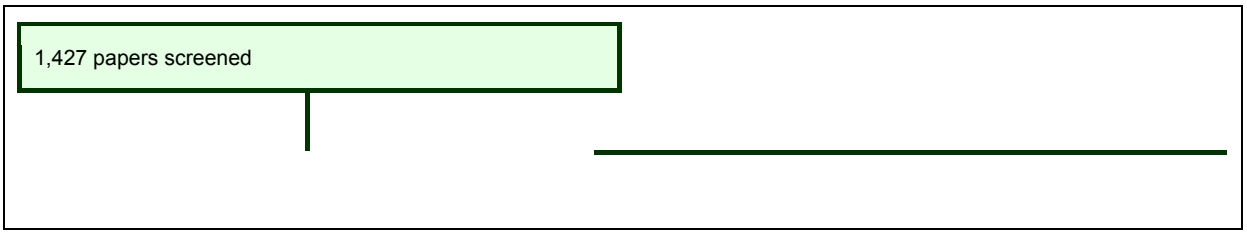
Synthesis of extracted evidence - Narrative synthesis, supported by the data extraction tables, was used to summarise the evidence base.

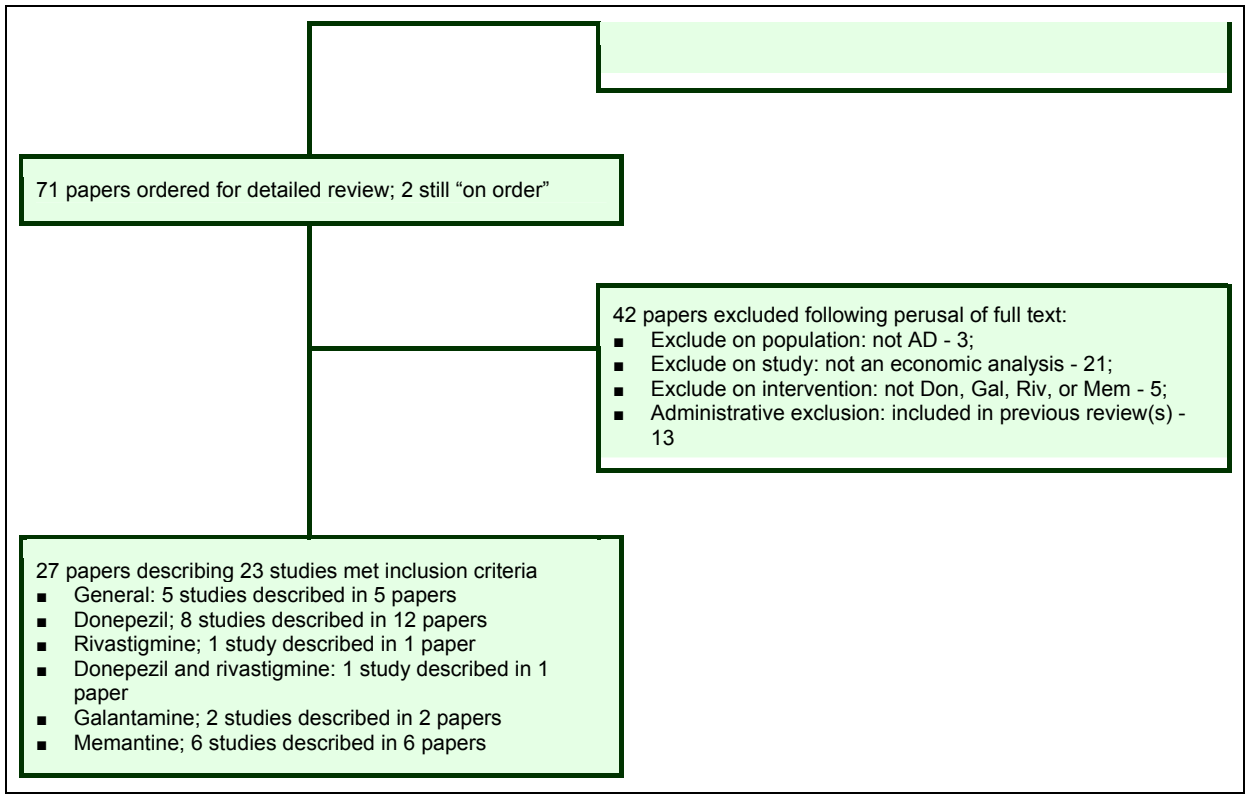
5.2.2. Results

The flow of papers is summarised in *Figure 56*. In brief over 1,400 citations were identified from the searches, 71 of which were ordered in full; 2 of these could not be retrieved but from the information in the tile seemed to offer a low probability of representing additional included studies. Of the 69 which were retrieved 42 were excluded. The most common reasons for exclusion were the paper was not an economic evaluation or the paper had been considered in the previous guidance. Further details and references for these excluded papers are available on request.

27 papers describing 23 studies were included, and are detailed in the following sections. No additional potentially includable studies were identified from checking of reference lists of included systematic reviews or manufacturer submissions.

FIGURE 56 *Flow diagram for search, retrieval and inclusion of articles in systematic review of evidence on the economic evaluations of AChEis and memantine for treatment of AD*





5.2.2.1. Results - general papers not specific to one particular drug

There were 5 studies in this category [148-152].

There were 2 systematic reviews of economic evaluations. Green [148] offered a slightly extended search period to the systematic review of cost-effectiveness studies presented in the last TAR report, searching to the end of 2005. Two additional studies were identified in the period 2004 to 2005, both of which are included in this review. Oremus [151] searched to the end of 2007. He identified 11 studies published in the period from 2004 to the end of his review. 4 of these were included in the original TAR report and the remaining 7 are included in this review. Both systematic reviews concluded that further research on the cost-effectiveness of AD was the priority. Green suggested the need for improved model structures and model parameters, and Oremus the need for more economic evaluation alongside RCTs. A further systematic review addressed cost studies, subdividing these by country in which the estimate was derived [150]. All the UK-based studies had already been captured and discussed in detail in the last TAR report. The review reinforced the highly variable nature of costs from country to country.

Two further papers provided information on cost-effectiveness of AD medication. Although these were not considered of sufficiently direct relevance to be appraised in detail for this guidance they are noted because of potential relevance to future guidance. The first paper by Jones [149] is a randomised trial protocol for "DOMINO-AD" (donepezil and memantine in moderate-to-severe AD)[ISRCTN49545035]. It addresses the question of what treatment course should be pursued in the face of advancing AD in patients already receiving an AChEI like donepezil. It will test the effectiveness and cost-effectiveness of alternative approaches, such as continuing or stopping donepezil and/or starting memantine. Economic evaluation will be conducted alongside the trial. The second paper [152] is a Monte–Carlo model assessing the cost-benefit of screening for, and then applying unspecified anti-AD interventions including drug treatments. A further abstract on the use of donepezil in this context was also found and is mentioned in the relevant sub-section below [153].

5.2.2.2. Results – donepezil

There were nine included studies addressing the cost-effectiveness of donepezil [153-162].

Three further included papers were interim reports of [163;164] or correspondence on [165] the main study by Feldman [154]. One further included paper [166] was a conference abstract of the study by Teipel [162]. This contained very little information and was thus not considered further.

Finally as already mentioned in the general section a further included study reports an on-going economic evaluation alongside a trial investigating the cost-effectiveness of stopping donepezil, continuing donepezil, continuing donepezil with the addition of memantine, or stopping donepezil and starting memantine in AD patients who have stopped responding to donepezil [149].

The general features of the nine main included cost-effectiveness studies for donepezil are set out in *Table 74*.

TABLE 74 *Included economic evaluations of donepezil*

	Feldman 2004a[154;163-165]	Fuh 2008[155]	Getsios 2009a[156];[157]	Getsios 2009b[153]	Lopez-Bastida 2009[158]	Lu 2005[159]	Mesterton 2009[160]	Teipel 2007[162;166]
Publication type	Full paper	Full paper	Abstract	Abstract	Full paper	Full paper	Abstract	Full paper
Study purpose	Investigate costs to society of AD	Apply existing model toTaiwan	Enhance method of modelling cost-effectiveness	Assess cost-effectiveness of screening + treatment	Assess cost-effectiveness in Spain	Estimate impact on health care costs	Up-date estimates of cost-effectiveness using recently collected data	Apply existing model to Germany and extend time-frame
Country setting	Canada, France & Australia	Taiwan	UK	UK	Spain	USA	Sweden	Germany
Base year prices	1998 CDN$	2006 US$	2007 GB£	Not stated	2006 €	1999-2002 US$	Not stated; SEK	2004 €
Intervention/comparator	Donepezil 5-10mg for 24w vs placebo	Donepezil vs no pharmacological treatment	Donepezil 10mg vs standard care	Screen & treat with donepezil 10mg vs treat alone	Donepezil 5-10mg vs no drug treatment	Donepezil vs no donepezil	Donepezil for 1y vs placebo	Donepezil 10mg for 1y vs placebo
Study type	Economic evaluation along-side RCT	Markov model	Discrete event simulation model	Discrete event simulation model	Markov model	Case-control study	Markov model	Markov model
Model duration/cycle length	N/A	5y; 12m	10y; N/A	10y; N/A	0.5,1,1.5,2 and 2.5y; 1m	N/A	5y; 6m	5 and 10y; 12m
Number of states	N/A	4	N/A	N/A	4	N/A	12	5
Study group – AD	Moderate to severe AD; MMSE 5-17	Mild and moderate AD	Mild to moderately severe AD; MMSE10-26	Patients without AD>65 and having memory complaints are screened	Mild or moderate AD	ICD-9-CM diagnosis of dementia	Mild to moderate AD	Mild to moderate AD
Perspective	Societal	Societal	Health care system and societal	Health care system and societal	Health service and societal	Health care payer	Not stated; implied societal	Societal
Discount rate pa (costs/benefits)	N/A	3%;3%	3.5%; applied to "all outcomes" (i.e. both benefits and costs)	Not stated	3%;3%	N/A	Not stated	5%; unclear whether applied to both benefits and costs

	Feldman 2004a[154;163-165]	Fuh 2008[155]	Getsios 2009a[156]}[157]	Getsios 2009b[153]	Lopez-Bastida 2009[158]	Lu 2005[159]	Mesterton 2009[160]	Teipel 2007[162;166]
Industry role	Funded by manufacturers	Funded by manufacturer	Company employee listed as author	Company employee listed as author	Funded by Ministry of Health	Funded by manufacturer	Company employee listed as authors	German Centre of Gerontology. Statement of "no competing interests"
Study base-case "headline" findings	Donepezil cost-saving	Donepezil cost-saving from societal perspective	Donepezil highly cost-effective	Screening and treating is cost-effective	Donepezil cost-saving from societal perspective and cost effective from health service perspective for mild AD. Not cost-effective for moderate AD	Reduced health care costs associated with donepezil use in the Medicare manged plan studied	Donepezil cost-saving	Donepezil may be cost-effective but considerable uncertainties remain

Notes:
AD Alzheimers disease CDN Canadian
SEK Swedish kronor
N/A Not applicable
The details of the ninth included study for donepezil by Pattanaprateep are provided in the table for rivastigmine

The review of cost-effectiveness studies reveals some potentially valuable new evidence since the last guidance. The studies fall into four categories:

Primary economic evaluations: There were two studies in this category [154;159]. Of the two the study by Feldman is the more robust, representing a bottom-up costing alongside an RCT[167] in Canada, Australia and France in which 144 patients with moderate to severe AD (MMSE 5-17) were randomised to donepezil for 24 weeks and 146 to placebo. The societal cost per patient was CDN$ 9,904 in the donepezil group and CDN$ 10,236 in placebo, representing a net saving of CDN$ 332. When caregiver costs were excluded the cost was CDN$ 4,355 in the donepezil group and CDN$ 4,321 in placebo, representing a net increase of CDN$ 34 with donepezil. Conference abstracts[163, 164] present results for costs in the 145 patients with severe AD (MMSE 5-12) in which there was a net cost saving associated with donepezil of CDN$ 467. Thus, in turn this suggests that there would also be a societal cost-saving in the moderate group of AD (MMSE 12-17) of approximately CDN $ 200.

The study by Lu [159] was an observational study, and hence much more open to bias and confounding, but nonetheless also suggested that prescription of donepezil was associated with lower costs to a large Medicare managed healthcare plan. The difference in costs was US$ 2500 (95%CI 330-4671) and was adjusted for differences in a patient characteristics between cases and controls.

Both studies were funded by the manufacturer of donepezil.

Application of existing models to different settings: Three studies donepezil [155;158;162] essentially apply existing model structures to new settings, defined in terms of the health care systems in different countries. Parameters, where country specific estimates exist, were substituted for the parameters and assumptions in the parent models. The conclusions are consistent with the parent models which were reported in the last guidance indicating that donepezil is cost-saving, particularly when a societal perspective is employed. The study based in Germany [162] was perhaps more cautious in its conclusions than past models acknowledging the enormous impact of uncertainty on its cost-effectiveness estimates and also suggesting that implementation might not be justified in the context of the German reimbursement system. The Spanish study [158] was interesting in suggesting that cost-effectiveness might be better in mild AD, but not in moderate AD. This is the opposite conclusion to that reached by NICE in its last guidance.

In terms of industry involvement only the model by Fuh [155] was supported by the manufacturers. The models by Lopez-Bastida and Teipel [158;162] represent two of the few economic evaluations apparently performed independently of manufacturer influence.

Newly developed or up-dated models: Three conference abstracts representing two models donepezil [153;156;160] appear to represent novel approaches to modelling. The analysis of donepezil from the UK perspective has also been published as a full paper in early 2010.[157] In the model by Getsios [153;156;157] a discrete event simulation approach has been developed to deal with limitations of previous models. There is very limited information in the abstracts about the details of the model, but it seems clear from the full paper version[157] that the approach adopted is very similar or even identical to the manufacturer submission for donepezil for this NICE guidance. For this reason we did not explore it further at this stage, relying instead on the working model supplied by the manufacturer. The study by Mesterton [160] also provided very limited details to support the view that it genuinely provides and up-dated approach using new data on costs and utilities. Concerning results, both models in this category suggest that donepezil produces health benefit and is cost-saving, and so dominates the no drug treatment alternative. In the second abstract and the full paper using the Getsios model [153] the new model is applied to the question of whether screening for AD, followed by donepezil treatment is cost-effective relative to donepezil treatment in those presenting with AD. The screening approach is claimed to be cost-effective, although this is not an issue of direct interest in this appraisal.

Both models in this group of studies have been developed with the support of the manufacturer.

Other: There was one poorly described model [161] which claimed to have assessed the cost-effectiveness of donepezil, high dose rivastigmine and low dose rivastigmine relative to no drug treatment in a Thailand private hospital. The details were so scant however, that it is debatable whether the conclusions can be given any credibility.

5.2.2.3. Results – rivastigmine

There were only two included studies claiming to provide new evidence on the cost-effectiveness of rivastigmine [161;168]. Their details are summarised in *Table 75*.

TABLE 75 *Included economic evaluations of rivastigmine*

	Brennan 2007[168]	**Pattanaprateep 2005[161]**
Publication type	Abstract	Abstract
Study purpose	Assess cost-effectiveness of Exelon patch	Assess cost-effectiveness in Thailand
Country setting	UK	Thailand
Base year prices	Not stated; GB£	Not stated; Bhat & US$
Intervention/comparator	Exelon patch vs best supportive care	Donepezil 10mg vs high or low dose Rivastigmine vs no drug treatment
Study type	Model (type not stated)	Decision tree analysis
Model duration/cycle length	Not stated	Not stated; N/A
Number of states	Not stated	Not stated
Study group – AD	Moderate AD	Mild to moderate AD
Perspective	UK NHS	Health service
Discount rate pa (costs/benefits)	Not stated	Not stated
Industry role	Company employee listed as author	Not stated
Study base-case "headline" findings	Exelon patch cost-effective	Cost-effectiveness of high dose rivastigmine greater than donepezil, greater in turn than low dose rivastigmine

Notes
AD Alzheimers disease
N/A Not applicable

The first study[168] was a model which claimed to assess the cost-effectiveness of a rivastigmine patch. Unfortunately the scant methodological details undermine the credibility of its findings that the Exelon patch was cost-effective with a cost per quality-adjusted life year (QALY) of about £13,000 from a UK NHS perspective. The study had support from the manufacturer.

The second study [161] attempted to compare rivastigmine at high and low doses with donepezil and has already been described in the donepezil section. As already indicated the details of the modelling process are so scant that the credibility of the conclusion that high dose rivastigmine is more cost-effective than donepezil which is in turn more cost-effective than low dose rivastigmine must be questioned.

5.2.2.4. Results – galantamine

There were again only two included studies claiming to provide new evidence on the cost-effectiveness of galantamine [169;170]. Their details are summarised in *Table 76*.

TABLE 76 *Included economic evaluations of galantamine*

	Suh 2008[169]	**Suh 2009[170]**
Publication type	Full paper	Abstract
Study purpose	Assess clinical and economic benefits of galantamine	Apply existing model to Korean setting
Country setting	Korea	Korea
Base year prices	2002; KRW & US$	2007; US$
Intervention/comparator	Galantamine 8-24 mg/day vs community control over 1y	Galantamine vs usual care
Study type	Economic evaluation alongside controlled trial (non-randomised)	Markov model
Model duration/cycle length	N/A	5y; not stated
Number of states	N/A	3
Study group – AD	Mild to moderate AD (MMSE 10-22)	Mild to moderately severe AD
Perspective	Societal	Third-party payer
Discount rate pa (costs/benefits)	N/A	6%; 1.5%
Industry role	Sponsored by manufacturer	No financial support. Statement of no conflicts of interest
Study base-case "headline" findings	Galantamine is cost-saving	Galantamine is cost-effective relative to usual care

Notes
AD Alzheimers disease
KRW Korea Won
N/A Not applicable

Both studies were by Suh [169;170]. The first [169] was an industry-sponsored economic evaluation alongside a controlled trial in which the costs of galantamine administered in the context of an RCT comparing different galantamine doses was compared with the costs in a community derived untreated control group. The duration of the study was one year and showed a cost saving of US$ 5,372.

The second study by Suh[170] is an economic model, in which an existing framework is applied to the Korean setting. The results suggest that from the perspective of a third party payer over five years galantamine is cost effective relative to usual care, cost per QALY US$ 4939. The author claims that there are no conflicts to declare, but this is somewhat inconsistent with the manufacturer sponsorship of the previously mentioned economic evaluation alongside the RCT [169] in which the same author is the lead.

5.2.2.5. Results – memantine

There were 6 main included studies addressing the cost-effectiveness of memantine [171-176].

In addition, as already mentioned in the general section a further included study reports an on-going economic evaluation alongside a trial investigating the cost-effectiveness of starting memantine, with or without donepezil in AD patients who have stopped responding to donepezil[149].

The general features of the main included studies for memantine are recorded in *Table 77*.

TABLE 77 *Included economic evaluations of memantine*

	Antonanzas 2006[171]	**Gagnon 2007[172]**	**Guilhaume 2005[173]**	**Jonsson 2006[174]**	**Toumi 2009[175]**	**Weycker 2007[176]**
Publication type	Full paper	Full paper	Abstract	Full paper	Abstract	Full paper
Study purpose	Apply existing model to Spanish setting	Apply existing model to Canadian health care setting	Model validation	Apply existing model to Sweden	Model using up-dated predictive equations for time to FTC	Consider cost-effectiveness of memantine added to donepezil
Country setting	Spain	Canada	UK	Sweden	Norway	USA
Base year prices	2005 €	2005 CDN$	N/A	2004 SEK	2008 € & NOK	2005 US$
Intervention/comparator	Memantine vs standard care	Memantine vs standard care	Memantine vs standard care	Memantine vs no pharmacological treatment	Memantine vs standard care	Memantine + donepezil vs donepezil
Study type	Markov model	Markov model	Modelled outputs compared with actual outputs	Markov model	Markov model	Microsimulation model
Model duration/cycle length	2y; 6m	2y; 6m	Unclear	5y; 6m	5y; not stated	Life-time; not stated
Number of states	7	5	Not stated	13	3	N/A
Study group – AD	Moderately severe and severe AD	Moderate-to-severe AD; MMSE <19	MMSE<14	Moderately severe and severe AD	Moderate to severe AD	Moderate-to-severe AD; MMSE 5-14
Perspective	Societal	Societal	Not stated	Swedish public health care payer	Societal	Societal
Discount rate pa (costs/benefits)	6%;6%	5% (unclear whether applied to both costs & benefits)	Not stated	3%; unclear whether also applied to benefits	3%;3%	3%;3%
Industry role	Company employees listed as authors	Company employees listed as authors	Company employees listed as authors	Supported by unrestricted grant from company	Company employees listed as authors	Unclear
Study base-case "headline" findings	Memantine cost-saving	Health benefits with no additional costs	Modelled and actual disease course similar over 18m	Health benefits achieved with cost saving	Higher benefits with no additional costs	Improved clinical outcomes with reduced costs of health care

Notes
FTC full time care
CDN Canadian
N/A Not applicable
SEK Swedish kronor
NOK Norwegian kronor

Half of the papers [171;172;174] focused on the application of the analytic approach used in the previous guidance to different settings, and were thus thought unlikely to provide estimates of cost-effectiveness which responded to the criticisms raised in the last guidance. This was compounded by the likelihood that the analyses were not independent. All papers repeated

the conclusion put forward by the manufacturer at the time of the last guidance that benefits were achieved at reduced cost.

The study by Guilhaume[173] was of interest in providing reassurance that extrapolation of natural history by the Markov model corresponded with actually observed states, but was limited by the small amount of information available in the abstract. The study by Tuomi[175] appeared to represent a new approach to estimating the cost-effectiveness of memantine relative to standard care, but was again limited by the small amount of information available in the abstract. Normally we would have pursued additional information, but did not do so in this case because the modelling approach appeared similar to that adopted in the industry submission. This has been appraised in detail in a later section. The study by Weycker[176] also appeared to offer an updated approach relative to those encountered in the last guidance, and did not have an obvious connection with the manufacturer. It did, however, address a question not directly relevant to the decision question of interest.

Summary

The systematic review of cost-effectiveness studies published since the last guidance raises the following key points:

There have been further publications on cost-effectiveness of pharmacological interventions for AD in the general medical literature

These are generally supportive of the cost-effectiveness of the acetylcholinesterase inhibitors (donepezil in particular) and memantine in the treatment of AD at all stages of disease. Most work is supported by the manufacturers as it was in the last appraisal. There are however a few more examples of independent assessments[158;162] which although more cautious also support the cost-effectiveness of drug treatments for AD.

Many studies apply existing models to new settings and as such appear to add little further general understanding concerning the cost-effectiveness of AD drug treatments outside the new setting considered

There are some new economic evaluations alongside trials and other studies which appear to offer new evidence[154;159;169]. They support the cost-effectiveness of donepezil and

memantine, in contrast to the AD2000 study in the last guidance, but are all manufacturer supported.

There also appear to be a small number of novel approaches to modelling, attempting to overcome problems observed with previous models. The most obvious of these is the discrete event simulation model of the cost-effectiveness of donepezil[156]. This will be considered in closer detail as part of the assessment of the manufacturer's submission.

6. Assessment of industry submissions to NICE

6.1. Introduction

Four manufacturer submissions were potentially available for this MTA. However, Novartis did not submit an economic evaluation and Shire only provided a critique of aspects of the previous SHTAC model which they felt remained unaddressed. The remaining two manufacturers both submitted economic evaluations based on decision models, and they are both critiqued in this chapter. A critique of their clinical effectiveness evidence reviews can be found in Section 4.4.

6.1.1. Decision Support Unit Involvement

The decision support unit (DSU) was asked to examine the technical accuracy of the Eiasi/Pfizer economic evaluation for donepezil, as it was produced using software (ARENA) which the TAG were unfamiliar with. According to Section 3.2.10 of the Guide to the MTA process, it says that models should be submitted in standard software, and if manufacturers plan to submit models in non-standard software prior agreement should by sought. Pfizer requested that their model be submitted in ARENA software which NICE accepted on the basis that training would be provided to the Assessment Group. Although some training was provided further expertise in ARENA software was required, thus, the DSU were asked to help complete this task. The DSU report has been fully integrated into this chapter. The DSU did not examine Lundbeck's economic model.

6.2. Lundbeck (memantine) – Critique of economic submission

6.2.1. The decision problem

The manufacturer of memantine submitted a model-based economic evaluation comparing it with no pharmacological treatment. The model is based on a Markov approach and health outcomes were expressed as QALYs. A NHS and PSS cost perspective was used and

future health effects and costs were both discounted at 3.5% per annum. The patient cohort consists of individuals with moderate to severe AD as measured using a number of functional and behavioural instruments, but not MMSE. The model was run for two patient populations, with the starting characteristics shown in *Table 80:* for 1) a 'general moderate to severe AD' group and 2) a group considered to have baseline symptoms of agitation, aggression or psychosis as defined as a score of ≥3 on the NPI (referred to as the APS sub-group). The APS subgroup was included because the manufacture believes there is evidence that treatments are particularly effective in this group. A similar argument was put forward in Lundbeck's submission in the previous appraisal, although the Appraisal Committee was critical of the 'overly broad' way the sub-group had been defined, an issue that was also raised at the Appeal hearing). This point is acknowledged in Lundbeck's current submission.

6.2.2. An overview of how the model works

All hypothetical patients enter a health state, termed 'pre full-time care' (pre-FTC). All individuals are assumed to have moderate AD, as defined in *Table 80.*

Three health states are defined: pre full-time care (pre-FTC), FTC and death. The model cycles monthly over 5 years. This structure is in line with the AHEAD model, used in the previous appraisal, although memantine was not evaluated using it, although the definition of FTC varies. All individuals enter the model in the pre-FTC health states. Patients who receive memantine do so at the beginning of the model and remain on it all times unless they enter the FTC health state or die. The baseline (no treatment arm) probability of moving between the pre-FTC and FTC health states was assessed using a risk equation, derived from a non-controlled longitudinal UK-based prevalent cohort study (the LASER study) of people with Alzheimer's Disease (AD). The probability of death was also derived from this source, estimated using a Weibull function – the same equation is applied to both pre-FTC and FTC health states, meaning that the probability of death does not change with increasingly progressive disease. Weighted mean differences derived from a meta-analysis of 6 RCTs are applied to the risk equation as a method of incorporating memantine's treatment effect Utilities were estimated using a mapping exercise and data from the LASER study (n=98) relating to people with moderate to severe AD.

The model is run probabilistically, although not all of the appropriate variables are specified as distributions. Memantine was predicted to be less costly and more effective than standard

treatment in the base-case for both patient groups (*Table 78*). The base-case cost-effectiveness acceptability curves are not shown in the submission, but generated directly from the model programming and taken at face value, suggest that the probability of memantine being cost-effective is greater than 90% for both sub-groups at all willingness to pay for an additional QALY.

TABLE 78 *Baseline probabilistic results taken from the MS*

	Cost £ (2009)	QALYs	ICER*
General group			
Memantine	92.971	1.534	Dominant
Standard care	94,687	1.503	
APS sub-group			
Memantine	93,663	1.566	Dominant
Standard care	98,639	1.496	

* ICER, incremental cost-effectiveness ratio; Dominant indicates a treatment is more effective and less costly than the comparator

6.2.3. Comparator treatment options

The model compares memantine with no pharmacological treatment. This comparison is partly appropriate since NICE's current guidance does not recommend the use of memantine in moderately-severe to severe patients, and it is the only product to have marketing authorisation for individuals with relatively severe disease. However, the marketing authorisation for memantine has changed since the previous appraisal. It is now licensed for people with moderate to severe AD. Thus, in theory the AChEIs are also now appropriate comparator technologies at a moderate disease stage. Note however, that no RCTs directly comparing memantine and AChEIs monotherapies have been reported.

6.2.4. The risk equation – estimating the monthly probability of entering full-time care

One of, if not the key, element(s) to the model is the risk equation used to estimate the monthly probability of moving to the FTC health state. The risk equation was derived using a sub section of patients from the LASER-AD study. The LASER study included a total of 224 individuals at various stages of disease. This particular analysis was restricted to 117 (52%) of individuals, as the remaining 107 were already considered to require FTC at the time of enrolment. A statistical model was developed using the corresponding data set to estimate time dependent probabilities of moving between the pre-FTC and FTC health states based

on a number of patient characteristics and time. The final baseline equation is shown in *Table 79* and the baseline starting characteristics of the two populations are show in *Table 80.*

TABLE 79 *Baseline equation (p268 of MS)*

Variable	Coefficient	SE	p>Izl
Ln(time in month)	3.3195	0.4965	<0.001
Baseline ADAS cog total score	0.0330	0.0147	0.0247
Baseline ADCS-ADL total score	-0.0877	0.0164	<0.001
Baseline NPI total score	0.0377	0.0154	0.0140
ADAS-cog total score (slope)	0.8122	0.2798	0.0037
ADCS-ADL total score (slope)	-2.4072	-0.3995	<0.001
Intercept	-11.1343	1.8284	<0.001

TABLE 80 *Memantine model patient characteristics (p 269 of MS)*

Parameter	Mean	SD*
General population		
ADAS-cog baseline	36.30	1.70
ADCS-ADL baseline	45.00	1.87
NPI baseline	18.54	1.86
ADAS-cog slope	0.6116	0.0809
ADCS-ADL slope	-0.7503	0.0876
APS sub-population		
ADAS-cog baseline	40.30	2.66
ADCS-ADL baseline	45.60	2.31
NPI baseline	22.45	2.21
ADAS-cog slope	0.6179	0.1216
ADCS-ADL slope	-0.7775	0.1157

*used in the probabilistic analysis

More details of how the equation was derived are supplied in Appendix O of the submission, marked academic in confidence. In terms of the data collection exercise, the model was said

to be based on data collected at months 6, 18, 30, 42 and 54 months after baseline. The methods also state that exact dates when FTC was required were unknown, only changes in FTC requirements at the above corresponding time points, meaning the data were analysed using discrete grouped data methods rather than a continuous time model.

The TAG has the following concerns with the derivation and use of this risk equation, although they are not listed in any particular order of importance. It is unclear how representative the patient sample is with respects to the general moderate to severe AD population in the UK.

Approximately two-thirds of the LASER-AD patients were receiving AChEIs and any related treatment effect does not seem to have been taken into account when constructing the equation.

FTC was defined as either entering an 'institution' or when individuals were considered to be 'dependent' in terms of requiring FTC from others. While the latter assessment was said to be based on domains on the ADCS-ADL (basic activities, domestic activities and communication), the details of this categorisation process are unclear eg. the threshold value for requiring dependence. This is important, since a third of patients over the 54-months were classified as becoming 'dependent'. No sensitivity analysis was undertaken to test the robustness of the final risk model to alternative assumptions regarding the definition of dependence.

Fifty nine percent of patients whose details were used in this specific analysis were said to have mild AD at baseline, thus the sample used to derive the risk equation is not representative of the baseline decision problem (treatment for moderate to severe AD). While it is possible to hypothesise that exclusion of mild AD patients may lead to an increased time to institutionalisation and therefore greater relative treatments effects, it is not clear this is the case as the probability of entering FTC and different stages of disease and disease progression might not be constant or linear. As something of an indicator of this potential issue, it is worth noting that while 58% of patients in the risk equation study were classified as requiring FTC over the 54 month period, examination of the Markov trace for the general AD population showed that approximately 58% of patients in the general AD population model had moved to the FTC health state in the standard treatment arm by month 25. One-way sensitivity analysis undertaken by the TAG showed that if the probability of

death from both health states was set to 0 (to isolate the independent effects of the risk equation), 58% of patients moved to the FTC health state by approximately month 23, and by month 54, 99% of patients had entered FTC. Thus it is not clear that risk equation when employed in the model, accurately predicts the probability of requiring FTC in terms of being consistent with the source data.

The predictive equation has not be validated against an external data source, therefore the degree to which the results are generalisable is unclear

The programming of the statistical model is poorly described, meaning there is concern that it may not have been used appropriately. Specifically, in addition to the baseline ADAS-cog total score, baseline ADCS-ADL total score and NPI baseline score, the rate of change of ADAS-cog and ADCS-ADL were also significant predictors of time to FTC (the submissions refers to these variables as slope parameters (*Table 79*). These values were then multiplied by what is also referred to as mean ASDS-cog and ADCS-ADL slope scores *(Table 80).* This second set of variables were also said to have been derived from the LASER-AD study but 1) there is no explanation of the methods used to derive these values 2) what indeed these values represent. Examination of the basic risk equation described on page 268 of the full manufacturer's submission suggests they are likely to / could represent the natural progression of the variables over time. For example, the value of -0.7503 might represent the change in ADSC-ADL per time interval. However, the equation on page 268 also suggests that these variables should change over time, as they are specified to the jth time interval, but the programming in the model does not allow for these values to change. A more standard approach to applying risk equations in economic models is to multiply relevant coefficients by the current values on an outcome to predict the probability of a future event, and then to recalculate this probability every time the value of the underlying outcome changes. However, this basic approach does not appear to have been undertaken. An alternative approach to this would be to multiply the rate of change (ie. the slope) by time to assess over all change, as indeed the manufacturer has done in the pre-FTC utility function.

6.2.5. Estimating relative treatment effects

Treatment effects were added to the underlying equation (*Table 81*) using results from a meta-analysis of six RCTS (MRZ 9001-9605/1, MEM-MD-01, MEM-MD-02, 99679, MEM-MD-10 and MEM-MD-12). Specifically, changes on the ADAS-cog baseline, ADCS-ADL

baseline and NPI baseline scores were meta-analysed and literally added to the related baseline variables in the risk equation.

TABLE 81 *Memantine treatment effects (p34 main Lundbeck submission)*

Parameter	Treatment effect	SD
General population group		
Baseline ADAS cog total score	-1.54	0.31
Baseline ADCS-ADL total score	1.53	0.62
NPI baseline	-1.34*	0.93
APS subgroup		
Baseline ADAS cog total score	-2.08	0.59
Baseline ADCS-ADL total score	3.59	0.85
NPI baseline	-2.49*	1.65

* Note that these mean values are not statistically significant at the 95% level.

As with the derivation of the baseline risk equation, a number of criticisms can be levied at this meta-analysis. In three of the studies, patients were said to have mild to moderate AD. While the submission acknowledges this and states that these individuals were removed from the analysis, it is unclear how this was done.

The meta-analyses used to estimate base-case treatment effects were all based on observed case analysis, which compared with LOCF, are likely to generate larger estimates of treatment effect.

Only two of the six (Resiberg and van Dyck) compared studies that are strictly in accordance with the stated decision problem: memantine monotherapy compared with placebo alone. Concerns with respect to pooling the data for all six RCTs have already been raised in the clinical evidence section of this report.

A related issue is that ADAS-cog is not measured in Reisberg or van Dyck. Instead it is stated that SIB scores from the two studies were transformed into ADAS-cog scores using a linear regression model computed on data from the LASER-AD study data. No useful details of this transformation process are provided. However, one-way sensitivity analysis performed by the TAG showed that setting the mean ADAS-cog coefficient to 0 instead of -1.54 (therefore removing any treatment effect on this variable) did little to change the results the coefficient.

The Reisberg and van Dyck study measured functional status using the ADCS-ADL_{19} (scores ranging between 0-54), not the ADCS-ADL_{23} (scores ranging between 0-78), which is the version used in the evidence synthesis. The manufacturer states that scores from the shorter version were 'rescaled' into scores for the longer version. However, there is no discussion of the methods used to do this, or the possible errors this might introduce. One way sensitivity analysis conducted by the TAG suggests that the results are particularly sensitive to the ADCS-ADL_{23} component of the risk equation. For example, replacing the coefficient of 1.53 (*Table 81*) in the general AD population base-case with 0, increased the ICER to about £33,000 per QALY from being dominant. There are two further points to note on this issue. First, visual examination of the forest plots provided by Lundbeck suggests smaller mean effects are likely to have resulted if the Reisberg study was excluded from the meta-analysis. Second, the meta-analysis on ADCS-ADL_{19} results conducted by the TAG using LOCF analysis using week 24-28 data, showed marginally statistically significant results (WMD 1.408, p=0.044) meaning it is not all clear memantine monotherapy is associated with improvements in functioning.

Lastly, the results from the baseline risk equation analysis showed that the NPI hallucination score was a significant predictor of time to FTC, not the NPI total score. It is however unclear which of these variables was estimated in the meta-analysis, but it is most likely to be the latter. If this is true, there is a disjoint between the treatment effects estimated by the evidence synthesis and the underlying risk equation since the NPI total score was not found to independently predict outcome. It should also be noted that results from the TAG's own meta-analysis, when restricted to RCTs that included individuals with moderate to severe AD who either received memantine monotherapy or placebo showed a non statistically significant difference in NPI total score in favour of memantine (WMD -1.6; 95%CI -4.739 to 1.523). However, despite all this, basic one-way sensitivity analysis suggests that the base-case was not sensitive to different parameter values (setting the effect of memantine to 0 on the corresponding risk coefficient) had a negligible impact on the results.

6.2.6. The probability of death

The base-case probability of death was estimated using a sample of the LASER-AD population, and specified using a Weibull function, where the hazard is a function of increasing time, but no other independent variables. Specifically, patients who were not institutionalised or dependent at baseline were said to be included in the analysis. The same

Weibull function was applied to both pre-FTC and FTC health states, therefore the probability of death within the model was not considered to change with increasing severity of disease per se. The base-case general AD population analysis showed that 50% of people died in the no treatment arm at approximately month 60.

There were a number of specific concerns with this part of the model. A third of patients in the LASER-AD study had mild AD meaning that the function might over estimate survival in people with moderate to severe AD. One way sensitivity analysis undertaken by the TAG suggests that the results are very sensitive to this variable.

No justification was given for excluding people from the analysis who were already receiving. However, a crude one-way sensitivity analysis undertaken by the TAG suggests that the results were not sensitive to this the probability of death each month.

There is no evidence to suggest memantine increases patient survival. However, applying the same survival function to both health states effectively means that people who progress to FTC stay there for relatively long periods of time (and therefore are assigned relatively large costs) if it is otherwise believed that progressive disease as represented by being in FTC is associated with increased mortality. Put another way, benefits and reduced costs of effective treatment are modelled by keeping people out of the FTC health state for as long as possible. Thus if it is likely that people in FTC have more advanced disease, and more advanced disease is associated with higher mortality, then the model is likely to over estimate the cost-effectiveness of memantine.

6.2.7. Costs

Memantine treatment costs were said to be £2.16 per day in the manufacturer's submission regardless of dosage or pack size, but it is not clear this is the case. The March 2010 MIMS states that a 28 tablet 10mg pack costs £34.50. Thus, 20 mg per day is equal to (£34.50 / 28)*2 = £2.46. While one way sensitivity analysis by the TAG suggests that this increased cost had little bearing on the base-case cost-effectiveness results, clearly its importance will be magnified if other changes are simultaneously made to the model, such as lessening the effect of memantine. The manufacturer also included the cost of a psychiatrist at the start of memantine treatment (£126) and a GP monitoring cost of (£35) every six months.

The costs associated with pre-FTC and FTC were estimated using resource data from the LASER-AD study, and by combining this information with unit costs from the most recent PSSRU publication. Resource use data was said to have been collected using the Client Service Receipt Inventory (CSRI), by interviewing patients / their carers every three months. In effect, it appears that data have been retrospectively collected every three months by interview, thus there must be some concern about the accuracy of recalling information over this period of time. A similar criticism was raised in the previous assessment report. In general, the resource use study is poorly described. For example, little is said about how many people provided resource use data and how missing data were handled, Thus, it is difficult to assess the validity of the results.

The monthly pre-FTC and FTC were calculated to be £724 and £3,267 per month respectively (or £8,688 and £39,204 per year). The value of £3,367 is a weighted average of people who were considered to have received FTC in the community (£852 * n=29/98) and people who were considered to be institutionalised (£4,282 * n=69/98). The annual values used in the previous Assessment Groups economic model were £3,397 and £11,247 respectively. Thus, even without allowing for inflation in the latter, these estimates appear to be very different. One reason for the large discrepancy is that the industry submission appears to include the costs that are borne by individuals, rather than the state – an issue in the previous appraisal – but the percentage is not explicit.

Examination of the manufacturers costing exercise shows that the main difference in costs between the pre-FTC and FTC health states is the time individuals spent in 'day hospitalisation' and nursing homes. Specifically, an average of 0.63 days in pre-FTC and 0.87 (if individuals were community based) and 8.97 days (if they were in an institution) for people in FTC, using a cost of £281 per bed day. For individuals who were considered to be in institutionalised FTC, an extra cost of £1,760 per month was added to this amount. The table referring to the references for the unit costs of £281 hospital bed per day and £573 per week in an institution refer to other pages in the submission. However, referring to the other pages revealed no further details.

6.2.8. Utilities

Health benefits to individuals with AD were measured and valued within the analysis, but potential benefits to carers were not included. Patient utilities were estimated using results

from individual items on three different instruments mapped onto the EQ-5D five domain classification system (as direct EQ-5D scores were said to be absent). However, it should be noted that the EQ-5D has previously been directly used to estimate mean utility values for people with AD.

The mapping methods were considered by the TAG to be particularly poorly described, thus the values should be treated with some caution. For example, it is said that data relating a sample from the LASER-AD cohort were used, but the basic sample demographics are not reported. Indeed, many other important methodological issues are not discussed including: why the (unspecified) mapping approach was chosen, who did the mapping, why these particular instruments were chosen in the first instance or how different model specifications could lead to different results.

From the mapping exercise, a mean utility value for the FTC health state of 0.336 was derived (the equivalent value in the previous SHTAC base-case was 0.34). Values for patients in the pre-FTC health state were not set at a static amount, rather they were adjusted according to ADCS-ADL total score each cycle, using the results from a regression analysis. However, while the LASER-AD study was said to be the data source, few other details are provided. For example, basic sample demographics are not provided, the ADCS-ADL total score was said to be 'the strongest' predictor of utility, but it would be useful to understand the relationship between utility and other explanatory variables. Moreover, no assessment of goodness of fit is provided or whether alternative models would have better fitted the data.

Pre FTC utility=0.202+0.008 (baseline ADCS ADL total scores + ADCS ADL change * time in months) where the ADCS-ADL total score relates to the specific treatment strategy. Because in the base-case, the ADCS-ADL total score was assumed to be higher for memantine, this in effect means that memantine patients accumulate more QALYs per time period while in the pre-FTC health state compared with the standard care arm.

On investigation, it was discovered that this specification leads to some logical problems. For example, when time is 0, the pre-FTC utility score is 0.562, but when time is greater than 40 months, the predicted value is lower than the (mean of) 0.33 associated with FTC. Moreover, no justification is given for having utility levels based on a function of declining ADCS-ADL total score for one health state and a mean (fixed) value in the other.

6.2.9. Extra sensitivity analysis on the general population base-case

A number of additional deterministic sensitivity analyses were undertaken by the TAG, some of which have already been reported in the above text. In this section, a number of further analyses have been undertaken.

- Using previous the SHTAC costs and inflating to 2009 prices, using indices provided by Lundbeck, produces an ICER of about £20,000 per additional QALY.
- Simultaneously making the above change and changing the utility values so that the utility equation for pre-FTC also extends to include patients in the FTC health state, produces an ICER of approximately £30,600 per additional QALY.
- Changing to both the previous SHTAC inflated costs and original SHTAC utility values, produces an ICER of about £23,000 per additional QALY.
- Extending the time horizon to consider long periods of time had negligible difference on the results.

6.2.10. Summary of memantine model comments

The submitted economic evaluation of memantine was based on a three state Markov model, with many of the inputs relating to a UK-based (LASER) study. The base-case submitted analysis suggested that memantine generated more QALYs at lower cost compared with standard treatment for both a general population of individuals with severe to moderate AD and for individuals in an agitation / aggression / psychotic symptoms (APS) sub-group. The results were particularly sensitive to treatment effects as measured using the ADCS-ADL, as it both the monthly probability of entering FTC and utility values were conditional on it. However, the TAGs general view is that the base-case results should be treated with some caution – broadly speaking for the following main reasons.

The model is poorly described in many places. Particularly with respect to the derivation and implementation of the underlying risk equation, the methods used to derive the utility functions and to transform some outcome scores from one scale (from the RCTs) to other scales (which were specified in the risk equation). Many of the model inputs were derived from the LASER-AD study, but it is unclear how representative it is of the general AD

population, and whether appropriate sub-groups have been used for the various sub-studies. The results from the TAGs own systematic review of the memantine monotherapy RCTs compared with placebo shows almost no statistically significant advantage of using memantine, only on the CIBC+ which is not included in this model. Thus at a face level, it is difficult to believe that there is at least a 90% probability memantine is cost-effective at all willingness to pay as the results from this model suggest. Lastly, no attempt has been made to compare the cost-effectiveness of memantine with the AChEIs in individuals with moderate AD.

6.3. Eisai / Pfizer (donepezil) - Critique of economic submission

The manufacturer of donepezil submitted a model based economic evaluation, which was built in ARENA software and incorporated a Microsoft Excel data input sheet. Numerous other text files were also included that are created from the Excel input sheet by a VBA macro. This model was critiqued by the Decision Support Unit at the University of Sheffield.

6.3.1. The decision problem

The model evaluated the use of 10 mg daily of donepezil compared with 'no AChEI' treatment. All individuals were assumed to stop treatment at a MMSE of 10 if they had not already done so. No attempt was made to compare the relative effects of the three different AChEIs using mixed treatment methods or against memantine in individuals with moderate AD. The patient cohort consists of individuals with mild to moderate AD as measured using MMSE (mild MMSE 20-26, moderate MMSE 10-19). The model runs over a lifetime horizon (set in the base-case to 25 years). In the base-case, the model suggests that treatment is less costly and more effective than no treatment, for individuals with mild or moderate AD.

All health outcomes were expressed in terms of QALYs, where total expected QALYs are a summation of associated patient- and care-giver values. A NHS and Personal Social Services cost perspective was said to have been used in the base-case; although this is later acknowledged not to be strictly true in the submission. Future health effects and costs were both discounted at 3.5% per annum over a lifetime horizon.

6.3.2. Rationale for choice of modelling framework

The model is based on a DES approach. In a Markov type analysis, individuals move between a set of pre-defined mutually exclusive health states over a fixed unit of time according to a set of transition probabilities, thus they are often referred to as discrete time models. This is in contrast to DESs, where a set of possible events is defined (along with associated costs and health outcomes) but the time between each event is variable in a first-order sense (ie representing individual variability rather than parameter uncertainty). Thus, DESs estimate times between events, with the sum of these intervals typically representing total life-expectancy.

Both discrete-time and –event models are useful when treatment costs and benefits are likely to accrue over relatively long periods of times. However, the limitation with Markov models is that the probability of moving from one health state to another is typically not based on an individual's prior experiences. A further limitation with Markov type models is that they become inefficient and demanding in a programming- and data requirement-sense if multiple health outcomes are possible as increasingly more complex sets of health states are needed (for example changes on different AD scales that are considered to be important predictors of costs and health outcomes). DESs potentially overcome both these problems, thus it is considered to be an appropriate modelling approach in this AD context (note later comments in this chapter however, that suggest this model is not a DES in the truest sense).

6.3.3. An overview of how the model works

The ARENA model submitted by the manufacturer is a generic model which has a variety of other modules/logic which are not relevant for the current decision problem. For example, it includes a screening module, and an option for patients to restart treatment as well as having the provision to estimate costs/utilities of two additional drugs along with donepezil and no treatment. Notwithstanding the model's capability to perform different analyses, this critique focuses on the issues in the model that are directly related to the cost-effectiveness analysis of donepezil against no treatment.

6.3.3.1. How patients are selected

The model utilises a weighted sampling approach to sample the patients in the model from the trial population. The trial population consists of 826 trial patients and there is a provision

to select the patients based on different characteristics such as age, sex, MMSE, etc. The two main subgroups utilised are a mild patient group (221 patients with 20<=MMSE<=26) and a moderate patient group (542 patients with 10<=MMSE<20). The model utilises 1000 patients and these are sampled from the corresponding subgroup utilising a weighted approach i.e. if using a mild population, 1000 patients are sampled from the 221 mild patients and are assigned the characteristics of the corresponding trial patient. These characteristics include age, sex, race, MMSE, NPI, ADL, IADL, previous MMSE and the change in MMSE in the previous year, as well as other information such as whether they are on psychiatric medications, whether they are living with their primary caregiver and if so, the caregiver's age and gender. These characteristics are specific to the individual patients and are assigned to patients as attributes.

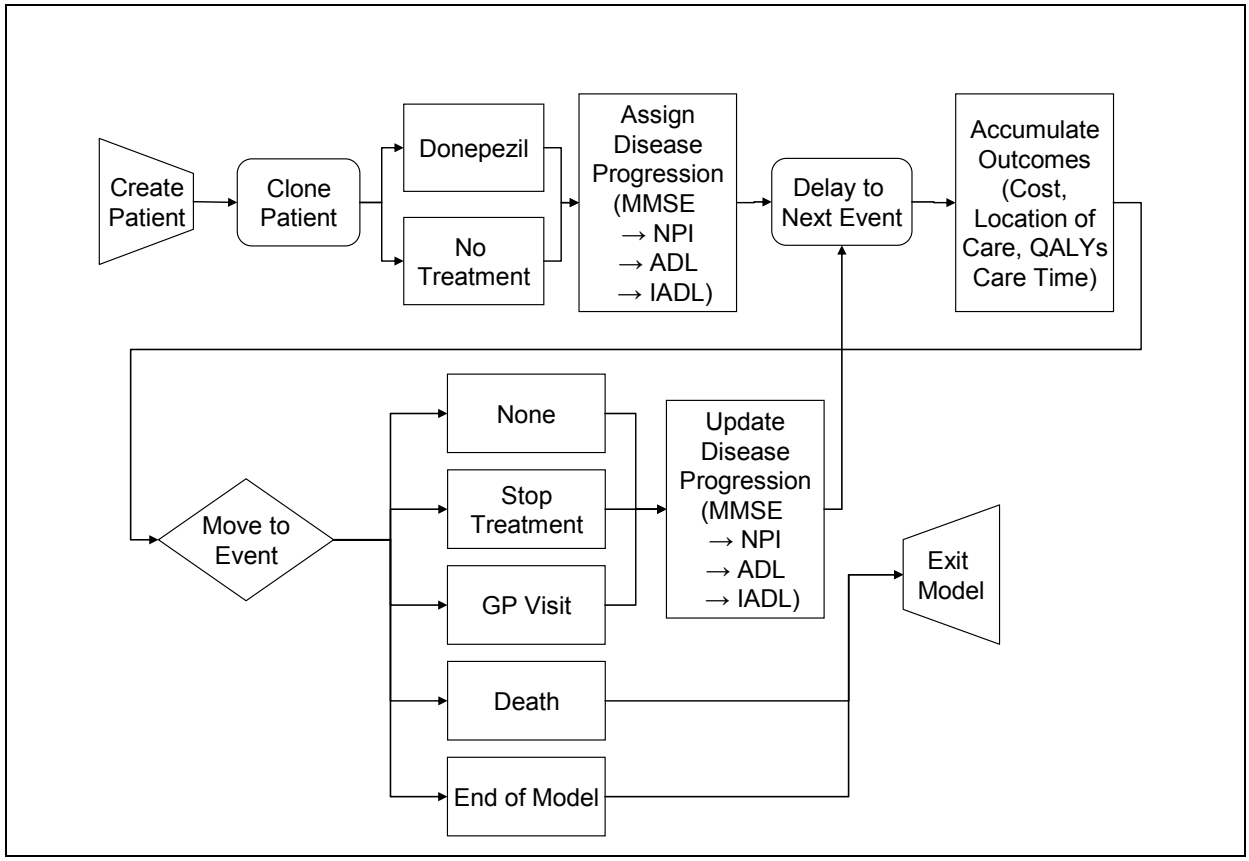
As there are fewer patients in the trial population, than in the sampled model population, it is likely that the same trial patient with be included more than once in the modelled population. As the sampling is weighted to achieve an age and sex distribution that is consistent with the UK AD population, this may mean that some patients whose characteristics are rare in the trial data set, but common in the UK AD population, may be sampled multiple times and their individual characteristics may have a disproportionate influence on the overall results.

The patients are then cloned i.e. each patient is separated into two identical patients with the exact same characteristics. One of the hypothetical patients is then allocated to the donepezil arm of the model and the other is allocated to the no treatment arm.

6.3.3.2. Model updates / disease progression

Disease progression is measured using a variety of outcome measures (referred to as attributes). The attributes of each patient are updated at different time intervals in order to replicate the progression of the disease and are then used by the model to perform cost effectiveness analysis. The model keeps track of four disease measures; MMSE, NPI, ADL and IADL (*Figure 57*). It should be noted that the MMSE equation uses annual increments while the other three equations use time as a continuous variable to estimate the new values. Annual changes in MMSE are first calculated, changes on the remaining three measures then follow, predicted by the change in MMSE as each of the other three equations includes current MMSE values as an individual term. Note that mortality is not dependent on choice of treatment.

FIGURE 57 *Simplified representation of the Alzheimer's disease model taken directly from the Eisai/Pfizer submission*



The underlying progression equations for MMSE are defined as follows;

Equation 5 (for the untreated cohort)

Annual change in MMSE = (- 5.4663 + norm(0,0.5) - 0.4299PM1 - 0.0042PM2 + 0.1415PM3 -0.0791PrevMMSEChange + 0.0747Ageorig)(Tnow-Tup)/365.25

Equation 6 (for the treated cohort)

MMSE on treatment = MMSE + (T_eff - 5.4663 + norm(0,0.5) - 0.429PM1 - 0.004PM2 + 0.1415PM3 -0.079PrevMMSEChange + 0.0747Ageorig)(Tnow-Tup)/365.2

where T_eff is 6.1583 if time is less than 20 weeks and 2.4671 otherwise. The treatment effects only last for one year after which it is assumed to be zero.

The underlying NPI equation in the model is defined as:

Equation 7

*NPI =(BaseNPI+5.74+norm(0,3.75)-0.64donepezil+0.03weeks-0.59NPIbaseline-0.0012NPI*weeks+0.24NPIrecent-1.74White-3.82Black+2.34psymed+0.12MMSEbaseline-0.22MMSErecent+)*1.44*

Where *psymed* is a dummy variable indicating whether or not individuals were receiving psychiatric medications, *black / white* are indicators of ethnic background and 1.44 is a scaling factor to convert the normalised scale of 0 to 100 scores to 0 to 144 (note that the reasoning / appropriateness of this transformation is not described in detail)

The underlying ADL equation in the model is defined as:

Equation 8

ADL =BaseADL+1.35+norm(0,2.48)-0.81donepezil+0.06weeks-0.79ADLbaseline+0.71ADLrecent+0.12MMSEbaseline+0.09age+0.81psymed-3.05Black-0.49MMSErecent

The underlying IADL equation in the model is defined as:

Equation 9

*IADL =BaseIADL+1.27+norm(0,1.9)+0.63donepezil+0.17weeks-(0.06Idonepezil*weeks)-0.84IADLbaseline-(0.002IADLbaseline*weeks)+0.84previousIADL-0.67male +0.20MMSEbaseline -0.28MMSErecent-0.16baselineADL+0.18ADLrecent*

The term norm(0,x) appearing in each of the disease progression equations is a random intercept parameter which is included to introduce patient level variation to the disease progression. This random variation is in addition to the variation provided by each patient having unique characteristics.

The patients are assigned a severity level based on their MMSE scores after every update. The severity categories and their MMSE ranges are shown in *Table 82.* The time spent in different severity levels are accumulated for all the patients in the donepezil arm as well as the no treatment arm. The proportion of patients in institutional care is dependent on the severity level are also as shown in *Table 82*.

TABLE 82 *Proportion of patients in institutional care according to severity level*

MMSE	Severity scale	Home	Institutional Care
25 to 30	Mild	87.1%	12.9%
20 to 24	Mild Moderate	74.4%	25.6%
15 to 19	Moderate	61.7%	38.3%
10 to 14	Moderate Severe	49.0%	51.0%
0 to 9	Severe	30.0%	70.0%

Even though the model utilises an individual patient approach, the patient and caregiver utilities are estimated using average values. For example, in the patient utility equation the "Institutionalised" covariate is, strictly speaking, a factor or dummy variable that takes a value of 1 if the patient is institutionalised and 0 if not. The cost effectiveness model does not classify individual patients as institutionalised or not. Rather they are assigned a probability of being institutionalised based on their MMSE score.

QALYs are estimated for both individuals with AD and a care giver. The utility functions are specified as follows:

Equation 10

$$Patient\ utility=0.408+0.01MMSE-0.004NPI—0.159Institution+0.051Living\ with\ care\ giver$$

Equation 11

$$Care\ giver\ utility=0.9-0.003Age\ caregiver+0.03Male\ care\ giver+0.001AgePatient+0.00MMSE-0.001NPI-0.001ADL-0.0004IADL+0.01Psymed$$

Where Psymed is a dummy variable indicating whether or not AD patients were receiving antipsychotic medications.

The costs for both donepezil treated and untreated patients are estimated by accumulating the treatment costs (for patients under treatment) and the patient care costs for home or institutional care. These monthly patient care costs are based on severity level as seen in *Table 83*. Again, although the model is based on an individual patient approach, patient care costs are estimated by multiplying the weighted averaging based on severity level by the time spent in that severity state. Drug treatment costs are accrued according to the number

of days on treatment. In addition to the drug treatment costs, patients also incur the cost of a medical consultation every 6 months whilst on treatment.

TABLE 83 *Monthly patient costs according to severity level and location of care*

Severity	Monthly Medical Costs (Home)	Monthly Medical Costs (Institutional)
Mild	£687	£2,801
Mild Moderate	£742	£2,801
Moderate	£798	£2,801
Moderate Severe	£878	£2,801
Severe	£957	£2,801

*PatientCareCosts = (Probability of home care*monthly home medical costs + Probability of Institutional care*Monthly Institutional Costs)*(Tnow-Tup)*12/365.25*

DrugTreatmentCosts = TmtCosts(Tnow-Tup)*

The caregiver times are estimated by the model but the caregiver costs are not taken into account. Hence, the total costs are calculated by adding the treatment costs to the patient care costs and the model estimates both discounted and undiscounted values of total costs. The discounted and undiscounted costs accumulated in different severity levels are also calculated.

6.3.3.3. Possible events

Patient characteristics are updated and the costs along with QALYs are calculated every time the patient undergoes an event. The events that occur in the life of a patient and the times when they occur are presented in *Table 84*.

TABLE 84 *Events occurring in the life of a person with AD*

Event	Time (in days)
Start treatment	0.01
Checks for discontinuation*	0.02, 91.3, 182.6 and 365.25
Regular updates	Every 3 months
Doctor visit	Every 6 months while in treatment
Stop treatment	Patient specific time
Death	Patient specific
Last update/model end	9131.25

* this event is used to assign a new Stop Treatment time

6.3.4. General concerns with the model and estimation of model inputs

6.3.4.1. The annual change in MMSE regression equation

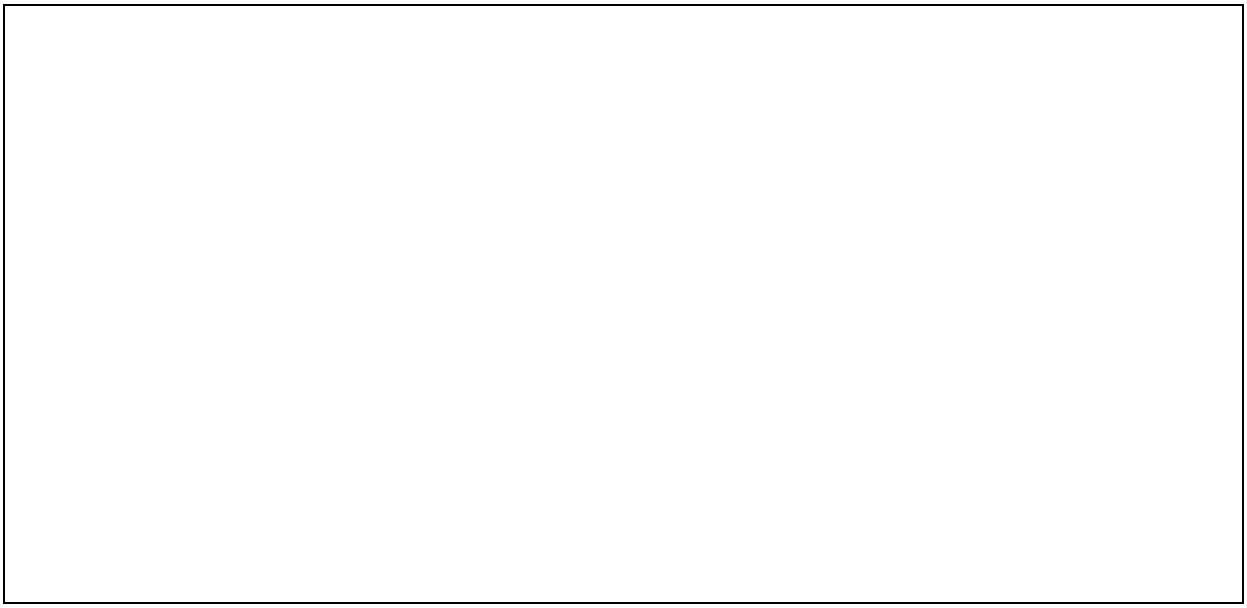
The MMSE regression equation has been derived from a sample of AD patients (721/1,094) from a US patient registry (the CERAD study); note that manufacturer's submission states that the registry only includes individuals who have never received treatment for their AD. While the principle of estimating a risk equation from a cohort study (and applying a treatment effect derived from RCT evidence) is considered to be sound, there are a number of concerns with the way the manufacturer has undertaken this analysis, meaning it is difficult to critique.

There is an overall lack of detail as to how the equation was constructed; Appendix J of the submission contains few additional details to the main submission. Specifically, the participants in the US CERAD study are not described in any detail, thus it is unclear how representative they are of UK individuals with mild to moderate AD. For example, it is stated on page 89 of the main submission that the CERAD data base does not include 'treated' patients. Little further discussion of this point is provided but it suggests that individuals included in the study might not necessarily be representative of a typical mild to moderate AD population. Additionally, corresponding model statistics, such as goodness of fit, are not provided and there has been no to attempt to validate the MMSE risk equation against external data sources, a point noted by the authors of the original economic model (Getsios).

In Appendix H of the submission, the manufacturer notes that the annual rate of change in MMSE was notably different when RCT data were used instead of individuals from the

CERAD study; this point is to some extent illustrated in *Figure 58* (Figure 4 in Appendix H of the submission). Specifically the submission states that using the alternative source of data led to 'no change or a small annual change in MMSE scores <20 and potentially large declines for those with values above 20'. A reason for this possible discrepancy is suggested - shorter measurement intervals in controlled studies - but it is uncertain that this in itself is sufficient justification for choosing one source over another, or whether it indeed suggests more reason to use it as the primary source. Furthermore, the patterns of change observed in CERAD were broadly more in line with what has been previously reported, as the trial data indicated an improvement in MMSE in some untreated patients *Figure 58.*

FIGURE 58 *Relationship between annual rate of change in MMSE and source data (taken from the Eisai / Pfizer submission)*



6.3.4.2. Changes in NPI, ADL and IADL scores

As already noted, three regression equations similar to the MMSE equation are used to predict the progression of NPI, ADL and IADL over time. For the NPI scale, data from four RCTs were used (not CERAD). The submission is not specific about the source of information used to estimate changes on the ADL and IADL scales. Indeed, while few methodological are details are provided one point of concern is that the ADL and IADL scales appear to be a composite of a number of different instruments although there is no discussion of how these transformations were undertaken, employed in terms of adding in

treatment effects or the errors this process might introduce. The following sentence has been taken from the MS:

> *"clinical trials measuring ADL and IADL used a variety of scales so 'standardised scales' were constructed using items from the various measures in order to link trial results to the utility function"*

6.3.4.3. Estimating treatment benefits

The effects of donepezil are included in the model through terms in the MMSE, NPI, ADL and IADL regression equations. All four equations include a direct donepezil treatment effect. The NPI, ADL and IADL equations also contained an MMSE term, meaning that changes in MMSE caused changes in the decline of these scores. More specifically, better maintenance of MMSE scores are predictive of a slower decline on the other disease measures. The IADL predictive equation also included an interaction term between donepezil treatment and duration of treatment; indicating that donepezil's effect increases over time.

The treatment effects on the NPI, ADL and IADL scales appear to have been estimated at the same time that baseline disease progression was estimated, as RCT was evidence used. However, the MS states that the terms used to estimate treatment effects of donepezil on MMSE were estimated using a 'similar' model to that derived for the baseline, as CERAD did not include treated patients. Appendix H of the submission suggests that results from eight RCTs were included in this analysis, but few other details are useful details of this 'similar' model are provided.

There is some concern that effects of donepezil have been double counted. Whilst treatment affects both MMSE, NPI, ADL and IADL directly as covariates in the four regression equations, there is also an additional link between the measures since MMSE is also a covariate in the NPI, ADL and IADL regression equations.

Treatment effects were assumed to be different after week 20, compared to weeks 21 to 52. However, the rationale for this cut off point is broadly stated to be 'after careful consideration of the data, and an attempt to maximise goodness of fit given insufficient data to consider to alternative functional forms' (page 89 of the main submission). Few other details are provided.

The MMSE treatment effect is modelled as an "absolute" benefit rather than using relative risk methodology. i.e. for each period on treatment there is a fixed, absolute change in MMSE. This assumes that all patients receive the same benefit, with a mean value derived from trial populations irrespective of their characteristics such as severity, age and sex. This absolute benefit is then applied to the untreated progression which is based on the CERAD data.

It is often assumed when building models that the relative risks from a trial are independent of the baseline risks and can therefore be applied to baseline risks estimated from cohorts which may be more representative of the population being treated. However, there is some concern about the approach taken in this instance as it is questionable that the reduction in progression achieved by treatment and estimated from the trial data is independent of the underlying rate of progression. MMSE treatment effect is one of the key drivers in this model, meaning that if the absolute treatment effect is not transferable from the trial patients to the CERAD cohort patients, the ICERs could be substantially different from those reported.

6.3.4.4. Patient utilities

Utilities are assigned within the model using an algorithm published in a Swedish study consisting of 208 from 272 people with AD and their carers, who were surveyed over a 12 month period. Utility values were measured using the EQ-5D and valued using a normative UK-tariff. A number of statistical models are presented, but the one used in the evaluation relates to data at all follow-up points, but is based on carer responses to the health status classification part of the EQ-5D. Note that age was not shown to independently predict utility values and the publication does not present statistical models based on patient responses.

The EQ-5D and associated valuation method were considered to be appropriate methods of assigning utility scores. However, concerns with the use of these algorithms included the following. MMSE score is used as an independent determinant of the change in baseline NPI score. However, both MMSE and NPI scores are used in the patient utility function meaning there is some concern that the effects of MMSE score have been double counted.

The published utility algorithm makes reference to the brief NPI. However, as most of the trials used the longer version of this instrument, the manufacturer converted the NPI coefficient to an alternative score (-0.018 became -0.004) – the transformation is poorly described and it is unclear whether it was appropriately undertaken.

The above patient utility function is based on proxy responses from AD carers, the publication does not show equivalent models based on AD patient responses. However, it is clear from within the publication that responses from individuals with AD and their carers were markedly different. Indeed non-adjusted results presented by MMSE strata suggest the choice of data set is likely to be an important determinant of utility (*Table 85*), an issue acknowledged by the original authors. For example, using the patient rated data set, there are few differences in utility scores between patients with mild and moderate disease. The differences are much more pronounced in the care giver related utility data set.

TABLE 85 *Patient EQ-5D utility values by MMSE strata from Jönsson et al 2006*

MMSE	26–30	21–25	15–20	10–15	9–0
Patient rated utility*	0.84	0.85	0.83	0.73	0.78
Care giver rated utility*	0.70	0.65	0.52	0.51	0.40

* n=649 data points

6.3.4.5. Care giver utilities

Care giver utilities were derived using SF-36 scores and the Brazier algorithm, using data from three clinical trials. The base-case results suggest that care giver QALYs contributed a much smaller amount to total QALYs compared with patient-related values (about a tenth depending on the exact scenario). However the TAG has some concerns with the final statistical model. For example, finer details of how the final utility equation was derived are not provided such as how the independent variables were chosen in the first instance or the overall goodness of fit. Moreover, the patient utility function suggests that entering an 'institution' significantly reduces AD patients' utility values. However, the care giver utility function does not include this term, when it is plausible to believe that such an event could increase carer's utility levels. It is unclear whether this was excluded because the relationship was not examined or because no such relationship was found. It is also worth noting that while ADL and IADL scales were shown to independently predict carer utility levels they are likely to incorporate broadly the same domains meaning there some reason to believe patient utilities are being doubled counted. The patient age coefficient is also

positive, suggesting care giver utilities increase with increasing patient age. Altogether, there is some concern over the robustness and underlying logic of the care giver utility function.

6.3.4.6. The probability of entering institutionalised full-time care and associated costs

A daily treatment cost of £3 was included for 10 mg donepezil, along with six monthly costs of a 'doctor' visit (£62.29 per visit, purported based on NHS reference costs for a geriatrician appointment; the assessment group could not confirm this unit cost from the source cited, and most consultant-led outpatient appointments cost from £100 to £170 in the national schedule of reference costs). However, the expected costs of care are by far the larger and more important costs components. Specifically the model includes the possibility individuals enter (institutionalised) full-time care (IFTC). As previously noted, the time to entering IFTC in this model is not an explicit event in a DES sense, rather it is modelled purely as a function of MMSE score. More specifically, individuals in one of five MMSE strata have a probability of either being cared for in the community or in IFTC, with increasing MMSE scores associated with a higher probability of being in IFTC (*Table 82*). The latter is also associated with higher costs (*Table 83*). Note that the costs of IFTC were not assumed to vary according to severity of disease.

The cost estimates in the industry submission are all taken from a report commissioned by the Alzheimer's Trust (Dementia UK report) in 2007, inflated to current prices. They include both health care and PSS costs but, unlike the original SHTAC model, no adjustments are made for the proportion of these costs for which the AD patient or their family is liable. Note that the costs of IFTC inputted into this model are approximately 3 times higher than those inputted into the original SHTAC model. Specific criticisms that can be levied at the accuracy and use of these cost estimates include the following. The costs are estimated on retrospectively collected resource use data for 114 individuals between January 1997 and June 1999. Thus, not only is the sample size arguably small, they may not represent contemporary standards of care. Similarly, unit costs have effectively been inflated from 1998 until the present year and are liable to similar criticisms. The authors of the report themselves state that care arrangements are likely to have changed during this time.

Disease severity in the report was classified as mild, moderate and severe disease using the Clinical Dementia Rating Scale. However, the model is divided into five severity groupings

dependent on MMSE score – no details are provided as to how the costs were divided up – but further investigation of these costs in the Dementia UK report indicate that Eisai have assumed that mild on the CDR scale is equivalent to MMSE>25, moderate on the CDR scale is equivalent to 15< MMSE <20, and severe on the CDR scale is equivalent to MMSE < 10. The two further categories of 19 < MMSE <26 and < 9 MMSE <16 are calculated as the means of the costs in their adjacent severity groups. I.e. the cost for 19 <MMSE<26 is the mean of the cost for 15< MMSE <20 and MMSE>25. Additionally, although perhaps not the largest concern, the resource use study was based on people with dementia rather than individuals with AD.

Information on the proportion of individuals living in the community and IFTC by severity of disease was estimated using a published report in 2007 of 445 individuals in a UK nursing homes (described in the study as being care homes for the elderly mentally infirm but excluding 'specialist' residences). However, this (important) component of the model is considered by the TAG to be particularly poorly described since the original report does not include specific statistics relating to the proportion of individuals who are living in the community or in institutionalised care. This issue is acknowledged in the manufacturer's submission however the assumptions and calculations used to generate the proportions in *Table 86* are lacking in any detail. This said, an obvious criticism of the use of this evidence is that the study was completed in individuals who were already in nursing homes – it did not include people who had not been admitted to care. Thus, it is difficult to understand how these proportions could be accurately derived from this data set in the first instance. Also note that a quarter of the study participants were estimated not to have dementia. The importance of this evidence as a driver for cost-effectiveness is discussed below.

6.3.5. General technical concerns with the model

6.3.5.1. Patient population

The modelled population is sampled from individual level data from three RCTs but it is weighting by age and sex to match the distribution of these variables in the UK AD population. The weighted sampling is done from the patient population after it has been filtered to include only mild or only moderate patients. It should therefore produce age and sex distributions that are similar in each severity category. However, the simulated moderate population has a better mean survival than the simulated mild population (4.603 v. 4.110

years). The manufacturer states, in section 3.4.14 of their submission, that this is because the simulated moderate population is younger and has a higher proportion of women. This may produce misleading results if patients with mild disease are actually more likely to be younger than patients with moderate disease in the UK AD population. It also suggests that the method used to weight the sampling to match the age and sex distribution in the UK is not functioning effectively. This could be because there are insufficient patients in the data set from which the population is sampled as previously discussed.

There is also some concern that other characteristics of the sampled population may not match the UK AD population, such as the likelihood of living with a carer, use of psychiatric medications, ethnicity, etc. For example, everyone lives with the carer in the sampled patient population, which is unlikely to be representative of UK patients. It has not been possible to investigate the sensitivity of the model to changes in the patient characteristics due to the way the model samples its patient population from the trial data.

6.3.5.2. Model structure

The DES approach has been used to track multiple patient characteristics, but these are updated at fixed intervals (e.g. 3mths). In a Markov model, a half-cycle correction would be applied to estimate the costs and QALYs based on the distribution of patients across the health states at the midpoint of each time-cycle. In this DES model, there is effectively a three month time-cycle but no equivalent "half-cycle type" correction is applied. Therefore if the time since the last update is 3 months, then the costs and utilities applied during those three months are based on patient variables at the end of the three months.

Even though it is claimed that this is a DES approach, the model calculates two of the most important parameters in determining costs and effects (patient care costs and utilities) using weighted averages in the same manner as a cohort model. Location of care (home or Institutionalised care) is not modelled on an individual level but is based on the mean rate for patients according to severity. The model is not a pure DES type model but incorporates elements of individual sampling and cohort modelling approaches.

6.3.5.3. Time to discontinuation of treatment

Different discontinuation rates are applied for different time periods within the model. The rates are presented in Table 8 of the manufacturer's submission as fixed probabilities over discrete time periods. In the model, it is assumed that the hazard is constant over each of these discrete time periods, allowing the hazard to be calculated from an exponential survival distribution. The hazard is then adjusted for three continuous risk factors which increase the risk of discontinuation. The individual's time to discontinuation, T_d, is then sampled using;

T_d= - *LN(UNIF(0,1)) / adjusted hazard.*

This time to discontinuation is re-sampled at the start of each discrete time period (0, 3, 6, 12mths). Each time a new sample is taken from the uniform distribution meaning that an individual who is sampled to have a higher than average risk of discontinuation in the first time interval (0 to 3 months) can then be sampled to have a lower than average risk of discontinuation in the next interval (3 to 6 months) even before the discontinuation risk has been adjusted to account for their individual risk factors. Using the same sample from the uniform distribution for each time interval would allow the risk of discontinuation to be estimated more consistently for the individual over the course of their lifetime, but still allow the hazard to be updated according to changes in their risk factor profile during the first year.

6.3.5.4. Error suppressions in calculations

There is an extensive use of various functions such as MIN, MAX, etc to suppress any implausible values that arise during calculations. For example, in utility and MMSE calculations (MX(0, utility)), MN(30,MMSE) and other similar expression are used to suppress any negative utilities values or any MMSE values greater than 30. It's the TAG / DSUs view that is would have been preferable for any implausible values predicted by the model to have been recorded as errors and investigated rather than being suppressed in the calculations.

6.3.5.5. Redundant programming syntax

The model submitted by the manufacturer is a generic model which has a variety of other modules/logic which are not relevant for the present Technology Appraisal. This redundancy is present throughout the model, which has hampered the review process. For example,

although the utility equation in the model is correctly implemented, it is defined as a combination of five different equations. This general lack of transparency means it is almost impossible to be certain that all the issues in the model have been identified.

6.3.6. Specific technical errors in the model

Life-expectancy

The manufacturer's submission states that expected survival was calculated by fitting functions of the form to the MRC CFAS data:

Survival (years) = A x (Percent Surviving) ^ B

The median survival estimates from the MRC CFAS data are given in Table 10 of the manufacturer's submission and the A and B parameters for men and women according to their age group are given in Table 11 of Appendix H to their submission. The model samples the life-expectancy of the patients as follows;

Time to death (in years) = A x UNIF(0,1) ^ B

where A and B are selected from Table 11 of Appendix H for the appropriate age and gender of the patient. The following mistakes in estimating the life expectancy of the patients in the model.

Male survival estimates are applied to women in one age category

For women aged 70 to 79 years, the expression (eTimeEvDeath) which is being used to select the appropriate A and B is referring to the data for men rather than women. The model is therefore underestimating survival in this group as median survival is greater for women in this age category. This error affects both treated and untreated patients.

No survival estimate for age 90.

The expression (eTimeEvDeath) which is used in the model to select the appropriate A and B values according to age and gender defines the oldest age category as age>90 rather than age>=90. It therefore does not generate an expected survival for patients aged 90. This effectively set the expected survival to zero for patients who start the model with age =90.

This error will essentially remove some patients from the model before they incur any costs or accrue any QALYs and again, it affects both treated and untreated patients. There are four patients aged 90 in the set of 826 trial patients from which the modelled population is sampled, but it is unclear how many times these patients are included within the sampled population.

MMSE Scaling

The PrevMMSEChange term used in estimating the updated MMSE (equation) is the annual rate of change and therefore the change since the last update has to be scaled to give an annual rate. This is calculated in the model as:

PrevMMSEChange=(MMSE - PrevMMSE)/(365.25/(TNOW-aLastUpdate))

However, it is our belief that it should be calculated as:

PrevMMSEChange=(MMSE - PrevMMSE)(365.25/(TNOW-aLastUpdate))*

Given that updates usually occur at three monthly intervals, the PrevMMSEChange scores is being underestimated by factor of sixteen.

Application of hazard ratios for discontinuation

The hazard for discontinuation of treatment is adjusted for three risk factors which increase the risk of discontinuation. These risk factors are baseline MMSE, current MMSE, and annualized change in MMSE. The hazard ratios for these risk factors are specified for different time periods during the first year. The hazard ratios for these risk factors are only applied during the first year of treatment and are then set to unity. In Appendix H, it is stated that a Cox regression model was used to estimate the hazard ratios. In a Cox regression model, the natural logarithm of the hazard is assumed to be a linear function of the form:

$$LN(hazard) = beta_0 + beta_1 x_1 + beta_2 x_2 + beta_3 x_3$$

Given that the risk factors included within the analysis are all continuous variables, it would be usual to present either the regression coefficient (beta) or the hazard ratios (HR) for an increase in one unit along the scale of the continuous variable (HR = exp(beta)). It can be seen from the equation above that the hazard ratio for a decrease in one unit is the

reciprocal of the hazard ratio for an increase in one unit. Likewise, the hazard ratio for an increase in two units is the square of the hazard ratio for an increase in one unit. More generally, if HR_1 is the hazard ratio for an increase in one unit from the reference range, then the hazard ratio for y units difference from the reference range is defined as follows:

$$HR_y = HR_1 \text{ ^ } y$$

In the model, the expressions aHRb, aHRc and aHRr are used to calculate the hazard ratio for the patient's baseline MMSE, current MMSE and annualised change in MMSE, as compared to the reference range for each of these variables. However, these are not being calculated in a manner which is consistent with a Cox proportional hazards model. Instead the following is being calculated:

HR_1 = *hazard ratio for 1 unit increase in MMSE*

For y > reference range: $HR_y = (HR_1\text{-}1)*y+1$

For y < reference range: $HR_y = (HR\text{-}1)*(1/y)+1$

TABLE 86 *Reference ranges used for the continuous risk factors*

Risk factor	Months 0–3	Months 3–6	Months 6–12	After 12 months
Baseline MMSE	18.8	18.8	18.8	1
Current MMSE	19.3	18.8	17.8	1
Annualised change in MMSE	4.31	-2.15	-2.69	1

Therefore, in the model MMSE scores which are lower and higher than the reference range both increase the risk of discontinuation rather than lower ones decreasing the risk and higher ones increasing the risk. The reference ranges used in the model are given in *Table 86* for information as these are not reported in the manufacturer's submission.

6.3.6.1. Discrepancies between the model and the submission

Five instances where the data in the manufacturer's submission does not match that being used in the model have been identified. The differences found were as follows;

- The constant in calculating the annual rate of decline in MMSE is -5.4663 in the model calculations instead of 5.4663 as mentioned on page 89 of the submission

- In the NPI equation, the coefficient for the interaction term, baseNPI x weeks, in the model is -0.0012 instead of -0.59 as reported in NICE submission (page 90 of the submission). The same coefficient is reported as 0.0012 in Appendix H, table 5
- The coefficient for the interaction term, baseIADL x weeks, in the IADL equation is -0.002 in the model instead of 0.002 as mentioned in table 7 of Appendix H
- The caregiver utility equation uses 0.013 as coefficient for PsyMed instead of -0.01 in the report (pg 93 of the submission) and in Table 15 of Appendix H
- The caregiver utility equation has a patient age term with a coefficient of 0.0014 in the model. Also, it has no term for patient gender as reported in page 93 of the submission

The first four discrepancies listed above were confirmed by the manufacturer to be typographical errors in the report and therefore do not alter the reported results. The fifth affects utilities of treated and untreated patients equally and therefore does not affect the incremental cost-effectiveness.

6.3.7. Probabilistic sensitivity analysis

The probabilistic sensitivity analysis (PSA) was replicated as detailed in the report i.e. 350 runs with 5000 patients each run for mild and moderate patients separately. Jack-knifing[177] has been performed on the results to identify the confidence intervals and they are reported in *Table 87*.

TABLE 87 *Jack-knifing analysis on manufacturer's model PSA (350 runs)*

	Deterministic			Stochastic		
	Cost £	**QALYs**	**ICER £**	**Cost £**	**QALYs**	**ICER £ (95% CI)**
Mild	-3,386	0.147	-22,975	-1,786	0.130	-13,764 (-18,873 to -8,768)
Moderate	-1,883	0.109	-17,310	-1,316	0.105	-12,585 (-17,727 to -7,553)

It can be observed that the deterministic mean is quite different to the stochastic mean even though all the ICER's indicate that donepezil dominates standard care. In fact, the mean cost savings and QALY gains are smaller in the PSA analysis for both mild and moderate patients which means that the base-case results presented in the submission are quite

optimistic relative to the probabilistic mean. This would suggest that the deterministic ICERs cannot be used as a good estimate of the expected cost-effectiveness and that the PSA analysis is the most appropriate to use. This said, there are concerns with the implementation of the PSA analysis in the model. In the health utility equations, all of the terms in the equation are varied within the PSA but each term is allowed to vary independently of the others removing any correlation between the terms. For the disease progression equations, only the intercept term and the treatment effects are varied within the PSA analysis. Again this removes any correlation between the intercept term and the other terms which are fixed. There are also specific concerns regarding the beta distributions used to describe the probability of institutional care as described below. The results of the PSA analysis should therefore be interpreted with caution.

Beta distributions for institutional care

The model uses beta distributions to describe the uncertainty in the proportion of patients receiving institutional care for each severity state. The alpha and beta parameters used to define the beta distribution are <1 for all severity states and are similar, but not exactly equivalent, to the average proportions in home and institutional care used in the deterministic analysis. When the alpha and beta parameters are both <1, this produces a U shaped beta distribution with asymptotes at 0 and 1 which does not seem to be a realistic distribution for this parameter. No details are provided on how the alpha and beta parameters, which are given in *Table 88* below, have been derived.

TABLE 88 *Beta distributions for institutional care used in the MS*

MMSE	Living in the community	Institutionalised	Distribution used in manufacturer model
Mild	87.1%	12.9%	Beta(0.86229,0.12771)
Mild-Moderate	74.4%	25.6%	Beta(0.73656, 0.25344)
Moderate	61.7%	38.3%	Beta(0.61083,0.37917)
Moderate-Severe	49.0%	51.0%	Beta(0.4851,0.5049)
Severe	30.0%	70.0%	Beta(0.297,0.693)

6.3.8. Amendments made to the base-case

Given the above concerns with the model, a number of corrections were attempted and additional sensitivity analyses were performed to examine the robustness of the results to alternative assumptions.

MMSE Scaling

The PrevMMSEChange term used in estimating the updated MMSE (equation) is estimated using a corrected scaling factor. The costs and QALYs estimated using the updated equation are presented against the base-case model in *Table 89* below based on a deterministic ICER after 20 runs with 1000 patients.

TABLE 89 *Cost effectiveness results compared with base-case model with corrected MMSE scaling*

	Base-case model			Base-case with corrected MMSE scaling		
	Cost £*	**QALYs***	**ICER £**	**New Cost £***	**New QALYs***	**New ICER £**
Mild	-3386	0.147	-22975	-2953	0.137	-21554
Moderate	-1883	0.109	-17310	-1612	0.102	-15813

Negative ICERs indicate donepezil is more effective and less costly compared with no treatment;

* indicates differences between treatment options

Life-expectancy

The expression eTimeEvDeath was changed to include patients aged 90 in the fourth age category and to select the appropriate estimates for A and B for women aged 70 to 79. The impact on results from these two combined changes is seen in *Table 90* based on a deterministic ICER after 20 runs with 1000 patients. The increase in cost-effectiveness of

donepezil can be attributed to using the correct life expectancy for women, which was underestimated in the base-case.

TABLE 90 *Cost effectiveness of base-case model with corrected life expectancy*

	Base-case model			Base-case with correct life expectancy		
	Cost £*	**QALYs***	**ICER £**	**New Cost £***	**New QALYs***	**New ICER £**
Mild	-3386	0.147	-22975	-4118	0.178	-23125
Moderate	-1883	0.109	-17310	-2022	0.117	-17296

* indicates differences between treatment options

Hazard calculations

Revised costs and QALYs were calculated by amending the expressions aHRb, aHRc and aHRr to use the correct method for calculating the hazard ratios as previously detailed. The effect on results from these changes is seen in *Table 91* below:

TABLE 91 *Cost effectiveness of base-case model with corrected hazard calculations*

	Base-case model			Base-case with correct hazard calculations		
	Cost £*	**QALYs***	**ICER £**	**New Cost £***	**New QALYs***	**New ICER £**
Mild	-3386	0.147	-22975	-3345	0.146	-22960
Moderate	-1883	0.109	-17310	-1922	0.110	-17417

* indicates differences between treatment options

New (deterministic) base-case results

The new base-case model is obtained by correcting the three errors identified in the manufacturer's model simultaneously. The model is 20 runs with 1000 patients (for both mild and moderate categories) and the ICERs are presented in *Table 92*. Note that the combined corrections have made little difference to the results.

TABLE 92 *Cost effectiveness of the new base-case model*

	Base-case model			New base-case		
	Cost £*	**QALYs***	**ICER £**	**New Cost £***	**New QALYs***	**New ICER £**
Mild	-3386	0.147	-22975	-3563	0.164	-21,713
Moderate	-1883	0.109	-17310	-1763	0.111	-15,824

* indicates differences between treatment options

New PSA results

More appropriate beta distributions relating to the probability of being institutionalised were entered into the model. However, as appropriate measures of variance were not available the following was undertaken, as outlined in Appendix H of the submission. "Where a standard error was not available, we used ±25% of the parameter mean to assign a 95% confidence interval and calculate the corresponding standard error estimate." Using this method, we have calculated the 95%CI for the proportion receiving institutional care and used these to derive alpha and beta parameters for the proportion receiving care at home as shown in *Table 93*.

TABLE 93 *Beta distribution for institutional care used in the DSU analysis*

MMSE	Living in the community	Institutionalised (95% CI)*	Distribution used in DSU analysis for proportion living in community
Mild	87.1%	12.9% (9.7% to 16.1%)	Beta(360.6,53.4)
Mild-Moderate	74.4%	25.6% (19.2% to 32.0%)	Beta(132.2,45.5)
Moderate	61.7%	38.3% (28.7% to 47.9%)	Beta(60.5,37.5)
Moderate-Severe	49.0%	51.0% (38.3% to 63.8%)	Beta(28.4,29.6)
Severe	30.0%	70.0% (52.5% to 87.5%)	Beta(7.6,17.7)

* calculated as proportion +/-25%

The model was then run 350 times for 5000 patients and jack-knifing was performed to calculate the confidence intervals. These results incorporate the revised beta functions in addition to the corrections made to the base-case to produce the deterministic results. It can be observed from *Table 94* that the confidence interval is smaller for the new base-case and this can be attributed to the fact that it uses the updated beta functions. The deterministic value is still towards the lower end of the interval obtained through the PSA analysis. These results incorporate the revised beta functions in addition to the corrections made to the base-case to produce the deterministic results.

TABLE 94 *Deterministic and PSA results for the manufacturer's base-case and new base-case with corrected Beta distributions (350 runs)*

	Deterministic			Stochastic		
	Cost £	**QALYs**	**ICER £**	**Cost £**	**QALYs**	**ICER £ (95% CI)**
Mild						
Base-case model	-3,386	0.147	-22,975	-1,786	0.130	-13,764 (-18,873 to -8,768)
New base-case model	-3,563	0.164	-21,713	-3,166	0.156	-20,282 (-22,837 to -17,730)
Moderate						
Base-case model	-1,883	0.109	-17,310	-1,316	0.105	-12,585 (-17,728 to -7,553)
New base-case model	-1,763	0.111	-15,824	-1,380	0.109	-12,678 (-15,309 to -10,057)

The model was also run 1000 times with 5000 patients to gather more accurate results and these are presented in *Table 95*. It can be observed that the confidence interval is smaller when using 1000 PSA samples rather than 350 PSA samples. Also, the stochastic mean is closer to the deterministic mean. These results suggest that it is necessary to run more than 350 samples to obtain an unbiased estimate using the PSA analysis.

TABLE 95 *Deterministic and PSA results for the new base-case with revised Beta distributions (350 & 1000 runs)*

	Cost £	**QALYs**	**ICER £ (95% CI)**
Mild			
Deterministic	-3,563	0.164	-21,725
PSA with 350 samples	-3,166	0.156	-20,282 (-22,837 to -17,730)
PSA with 1000 samples	-3,415	0.159	-21,433 (-22,354 to -20,515)
Moderate			
Deterministic	-1,763	0.111	-15,882
PSA with 350 samples	-1,380	0.109	-12,678 (-15,309 to -10,057)
PSA with 1000 samples	-1,703	0.111	-15,285 (-16,686 to -13,888)

6.3.9. Exploratory analyses on the new base-case

In addition to the above technical corrections, a number of exploratory sensitivity analysis were also run to examine the robustness of the results to alternative assumptions.

Proportion institutionalised

The proportion of patients institutionalised is dependent on the severity level as shown in *Table 82*. The patients are assigned a severity level based on their MMSE scores alone. As there was concern regarding the evidence used to estimate these proportions, further analyses were undertaken. Specifically, the assumption was made that disease severity levels (as measured using MMSE) has no effect on the probability of institutionalisation. This was implemented in the model by having the same proportion of patients institutionalised (36.5%) at all severity levels. This value of 36.5% is reported by the manufacturer as the overall percentage of institutionalised patients in the UK. The costs and QALYs are calculated and presented in *Table 96* below:

TABLE 96 *New base-case with fixed institutionalisation across severity levels*

	New base-case			New base-case with proportion of institutionalised patients set to 36.5% across all severity categories		
	Cost £*	**QALYs***	**ICER £**	**New Cost £***	**New QALYs***	**New ICER £**
Mild	-3563	0.164	-21713	2186	0.113	19,389
Moderate	-1763	0.111	-15824	1826	0.077	23,676

* indicates differences between treatment options

Assuming MMSE has no effect on institutionalisation, the ICERs for the mild and moderate populations have become £19,339 per QALY and £23,676 per additional QALY respectively.

Impact of institutionalisation on caregiver utility

As previously mentioned, caregiver utility is calculated using an equation which includes caregiver age and gender, the four main patient disease measures (MMSE, NPI, ADL, IADL) and use of psychiatric medicine. It does not contain any terms that relate to whether the carer is living with the patient and providing care in the home or whether the patient is living in an institution. The manufacturer's submission states that the caregiver utility equation has been derived using data from the Nordic, 324 and 312 trials, which are the same trials used to provide the patient data set from which the modelled population is sampled. Looking at this data set it would appear that all of the patients have the variable "living with patient" set to 1 suggesting that all patients had a caregiver living with them at the start of the study. They also state that information was not available on the impact of institutionalisation on caregiver utility. Therefore, caregiver utility is estimated in the model to be the same

regardless of whether the caregiver is living with the patient and regardless of whether the patient is receiving home care or institutional care.

Treatment reduces progression to more severe disease states which are associated with a higher risk of institutional care in the model. If institutional care is associated with an increase in carer utility due to a lower burden of care being placed on the primary caregiver, then reducing disease progression and lowering the average time spent in institutional care will reduce expected QALYs for the caregivers. The sensitivity of the model to alternative assumptions regarding caregiver utility was investigated by removing the utility decrement associated with NPI, ADL and IADL for patients receiving institutional care. This improved caregiver utility in both arms of the model, but the incremental effect of treatment on caregiver utility became negative as treatment delays institutionalisation which is associated with gains in care giver utility. The incremental costs and QALYs are calculated and presented in *Table 97*.

TABLE 97 *New base-case with modified caregiver utility*

	New base-case			New base-case with improved carer utility after institutionalisation			
	Cost £*	**Carer QALYs***	**Total QALYs***	**Cost £***	**Carer QALYs***	**Total QALYs***	**ICER £**
Mild	-3563	0.016	0.164	-3563	-0.010	0.138	-25,844
Moderate	-1763	0.011	0.111	-1763	-0.010	0.091	-19,399

* indicates differences between treatment options

Potential overestimation of treatment effect

As noted before, the the NPI, ADL and IADL expressions have an MMSE term as well as having a treatment benefit term. It is therefore possible that the effect of treatment is being overestimated. The importance of this structural assumption was investigated by using untreated MMSE values in the NPI, ADL and IADL progression equations for treated patients. The incremental costs and QALYs presented in *Table 98* show the same costs but with a reduction in QALYs as expected.

TABLE 98 *New base-case without MMSE effect on NPI, ADL and IADL*

	New base-case			New base-case without MMSE treatment effects carrying over into NPI, ADL and IADL		
	Cost £*	**QALYs***	**ICER £**	**New Cost £***	**New QALYs***	**New ICER £**
Mild	-3563	0.164	-21713	-3563	0.136	-26,130
Moderate	-1763	0.111	-15824	-1763	0.093	-19,001

* indicates differences between treatment options

Combined effect of the exploratory studies so far

This section presents the results of the new base-case model after making several changes to the assumptions to explore the combined effect. These were a) fixing the proportion of patients institutionalised across the severity levels; b) including the impact of institutionalisation on caregiver utility and c) removing the MMSE treatment effect from the NPI, ADL and IADL progression equations (*Table 99*).

TABLE 99 *New base-case with combined exploratory analysis*

	New base-case			New base-case: combined exploratory analysis		
	Cost £*	**QALYs***	**ICER £**	**New Cost £***	**New QALYs***	**New ICER £**
Mild	-3563	0.164	-21713	2186	0.085	25,831
Moderate	-1763	0.111	-15824	1,826	0.058	31,389

* indicates differences between treatment options

Regular update interval

Patient's disease status is updated regularly every three months and at these time points the costs and QALYs accrued since the last update are calculated. Updates are also made when other events occur such as stopping treatment or death but the timing of these events are unique to each patient. The new base-case model was run for different update intervals and the results are presented in *Table 100*. There seems to be a clear pattern, as the update period increases the cost savings and QALY benefits decrease and vice versa. However, we cannot be sure why the costs and QALYs vary in a systematic way in relation to the time period between updates. One possibility is that it may be due to the fact that the patient's attributes at the end of the period are applied to the whole period since the last update without any type of half cycle correction being used to reflect the fact that their attributes have been changing over that time period. If the patient's utility is falling over time,

this would systematically underestimate the QALYs accrued by the patient. If the patient's utility is falling faster in the untreated arm than in the treated arm, one would expect this error to overestimate the QALYs gained by treatment more for less frequent updates. Whilst here we see that the QALY gains are greater for more frequent updates. The cause of this behaviour has not been identified during our examination of the model and therefore we cannot exclude the possibility that it may be due to an error in the model logic.

TABLE 100 *New base-case with different update intervals*

	Mild Population		Moderate Population	
Update period	**Cost £**	**QALYs**	**Cost £**	**QALYs**
30 days	-4247	0.184	-2172	0.126
60 days	-3600	0.166	-1784	0.113
NewBasecase (90 days)	-3563	0.164	-1763	0.111
120 days	-2942	0.149	-1481	0.102

Distribution of sampled life-expectancy estimates

Samples of 5000 patients were generated using the distributions that are applied in the model (using Microsoft Office Excel 2007) and the summary parameters for these were compared with the MRC CFAS data. The median and interquartile ranges for each are presented in *Table 101*. The median survival estimates appear to match closely at older ages, but there are differences of up to 0.8 years in some age categories between the trial data and the sampled population which is being used to represent the distribution observed in the trial.

TABLE 101 *Median (and interquartile) survival estimates*

	MRC CFAS study (Table 10 Appendix H of MS)		5000 patients sampled from the distribution used in the model	
Age	**Women**	**Men**	**Women**	**Men**
65 to 69	7.5 (4.8-NA)	NA (9.1-NA)	8.1 (5.5 – 10.0)	11.8 (9.1 – 11.3)
70 to 79	5.8 (3.6-8.3)	4.6 (3.0-8.6)	6.0 (3.6 – 8.1)	5.4 (2.9 – 7.8)
80 to 89	4.4 (2.8-7.0)	3.7 (2.5-6.3)	4.8 (2.7 – 6.6)	4.2 (2.4 – 5.8)
>=90	3.9 (2.4-5.2)	3.4 (1.5-5.5)	3.9 (2.4 – 5.2)	3.4 (1.5 – 5.5)

The sensitivity of the model to differences in the survival estimates was investigated by running the model with the survival times fixed at the median and interquartile values taken from the MRC CFAS study (*Table 102*).

TABLE 102 *New base-case and new base-case with survival fixed at median, upper and lower interquartile range*

	Cost £	QALYs	ICER £
Mild			
New base-case	-3,563	0.164	-21,713
New base-case with survival fixed at median survival	-3,857	0.180	-21,395
New base-case with survival fixed at lower IQR	-2,669	0.129	-20,631
New base-case with survival fixed at upper IQR	-4,721	0.214	-22,102
Moderate			
New base-case	-1763	0.111	-15824
New base-case with survival fixed at median survival	-2,085	0.127	-16475
New base-case with survival fixed at lower IQR	-1,580	0.105	-15,056
New base-case with survival fixed at upper IQR	-2,239	0.133	-16,880

For males aged <70 years, no median or upper IQR are provided so it was assumed that the width of the interquartile range (IQR) from males ages 70 to 80 could be applied to estimate the median and upper IQR as 10.7 and 14.7 respectively. Additionally for women aged <70 years, no upper IQR is provided so it was again assumed that the width of the IQR from women aged 70 to 80 years could be applied to estimate the upper IQR as 10 years. The results show that whilst the cost-effectiveness estimate is sensitive to changes in the survival inputs, treatment still dominates no treatment even when applying the lower IQR for survival from the MRC CFAS study.

6.3.10. Summary of donepezil model comments

The version of the economic model submitted in the MS suggests that treatment with donepezil is less costly and more effective compared with no treatment, for individuals with mild or moderate AD. However, inspection of the manuscript and programming syntax suggests a number of important issues that should be considered alongside this claim.

Disease progression is modelled using four regression equations. First, changes in MMSE are predicted conditional on a number of independent variables (including treatment), followed by changes on NPI, ADL and IADL scales, also dependent on a number of variables (including treatment and current MMSE). However, there are a number of concerns with the appropriateness of the CERAD study used to estimate these equations and the possibility of double counting treatment effects, since MMSE was included as an independent term in the NPI, ADL and IADL scales. Moreover, the MS refers to the ADL / IADL scales as composite

measures without fully explaining how they were derived or how estimates of treatment effect measured using specific ADL / IADL scales in the various RCTs were linked to this equation.

The patient utility function includes a utility decrement if individuals enter institutionalised care. However, no such consideration is given to the possibility that a care-givers utility could increase at this time.

The model includes a probability that individuals at various MMSE strata require institutionalised care. However, on inspection the source data used to estimate these proportions only includes individuals who are already said to be in nursing homes. Thus it is unclear how these data have been used to estimate these proportions. Sensitivity analysis suggests that the base-case ICERs are sensitive to these proportions (see *Table 96*).

Closer inspection of the model also suggests that is not a pure DES approach but actually incorporates elements of individual sampling alongside some cohort modelling methods. In particular, it uses a cohort approach to estimate the costs of care and patient utilities based on the probability of institutionalisation rather than sampling the location of care for each patient.

The deterministic estimates of the ICER overestimated the cost-effectiveness of donepezil compared to the expected ICER obtained from the PSA analysis. This suggests that a robust PSA analysis is needed to determine the cost-effectiveness of donepezil. However, we also had significant concerns regarding the implementation of the PSA analysis and therefore the PSA results should be treated with caution.

A number of exploratory sensitivity analyses were undertaken to establish what the cost-effectiveness results would be if changes were made to some of the more important model assumptions, specifically where there was concern with respects to the quality of the inputted data. They included the relationship between MMSE and institutionalisation, the impact of institutionalisation on care giver utility and the potential overestimation of treatment effects that may be caused by the inclusion of the MMSE treatment effect within the NPI, ADL and IADL progression equations. Exploratory sensitivity analyses showed that the ICER for donepezil compared to no treatment could be as high as £26,000 per QALY in mild AD and £31,000 per QALY in moderate AD if alternative plausible assumptions are made for each of these key model assumptions.

Lastly, a number of technical errors within the ARENA programme were detected. While the corresponding corrections did not significantly alter the cost-effectiveness estimates or the implied decision, concerns remain that there may be further errors within the model as behaviour was identified which could not be explained when examining the use of an alternative update frequency. There are unresolved concerns regarding the way in which the model samples its population.

7. PenTAG cost–utility assessment

7.1. Defining the decision problem(s)

7.1.1. Interventions and comparators

The aim of this assessment is to review and update as necessary, NICE guidance to the NHS in England and Wales on the clinical and cost-effectiveness of donepezil, galantamine, rivastigmine, for mild to moderate Alzheimer's disease, and memantine, for moderate to severe Alzheimer's disease, which was issued in November 2006 and amended in September 2007 and August 2009.

Given the different licensed indications of the four drugs in the UK, for people with different levels of severity of Alzheimer's disease, this means that there are:

- Four alternative possible treatments/comparators for people with mild Alzheimer's disease (the three AChEIs plus best supportive care)
- Five alternative possible treatments/comparators for people with moderate Alzheimer's disease (i.e. the three AChEIs plus memantine plus best supportive care), and
- Two alternative possible treatments/comparators for people with severe Alzheimer's disease (i.e. memantine plus best supportive care)

Assuming that:

1. The three AChEIs should initially be treated as separate technologies (i.e. with different effectiveness estimates and different intervention costs), and,
2. That there may be sub-group evidence of their differential effectiveness for people with mild, moderate, or severe Alzheimer's disease (as defined by MMSE).

Then, there are, in theory 4 × 5 × 2 = 40, alternative technology adoption policies which might need to be modelled (i.e. accounting for all possible sequences of treatments across

the three levels of disease severity). This is clearly an impractical initial range of policy options to model, not least because the evidence of the effectiveness of the four drugs is unlikely to be available for all severity subgroups. Furthermore, evidence of the effectiveness for patients switching treatments, e.g. the effectiveness of switching from one AChEI to another when moving from mild to moderate disease, is also highly unlikely to exist in published trials. Ultimately, we found no published clinical effectiveness research which would support either of these potential modelling analyses.

Another new issue since the 2004 technology assessment's economic modelling, is that the range of disease severity treatable within the licences of the three AChEIs now overlaps with the severity range treatable with Memantine; patients with moderate Alzheimer's disease (MMSE 10-20). This means that, in theory, patients with mild Alzheimer's disease who progress to moderate Alzheimer's disease could now switch to or start on Memantine instead of continuing with their present treatment. Again, whether existing published trials would allow reliable estimates of the relative effectiveness of these treatment alternatives is doubtful (e.g. such estimates should ideally come from an RCT which had only recruited patients either diagnosed with moderate Alzheimer's disease or progressing to it, and allocates them to either memantine or one of the three AChEIs).

Therefore we have necessarily simplified our initial decision problem, and expanded it only when relevant research evidence was found which justified a more complex specification of the problem. This was necessary, for example, when the considerable difference in cost between patches and capsules for achieving the same daily dose of rivastigmine became apparent and needed to be reflected (see next section).

7.1.2. The decision problems to be modelled

Table 103 below shows the main alternatives that will be modelled in terms of the patient populations starting in the model and the treatment comparators. We have taken it as a given that the costs and outcomes (QALYs) to be estimated are those specified in the scope for this technology appraisal.

TABLE 103 *Populations and comparators to be modelled*

Decision problem	Simulated population	Starting comparators	Treatment continuation or switching
Decision problem 1a (Treating mild and moderate AD)	Existing (i.e. prevalent case) AD patients whose disease meets the eligibility criteria for receiving one of the 3 AChEIs (i.e. MMSE score-based ***mild or moderate AD*** (MMSE = 26 to 10)	Best supportive care (BSC) Donepezil Rivastigmine (×2)* Galantamine	Those who start on BSC, stay on it Drug treatment is continued until either (a) clinical decision to stop treatment (e.g. no longer responding) or (b) patient progresses to severe AD.
Decision problem 1b (Treating mild AD)	Existing (i.e. prevalent case) AD patients with ***mild AD*** (MMSE = 26 to 21)	Best supportive care (BSC) Donepezil Rivastigmine (×2)* Galantamine	As for Decision problem 1a
Decision problem 1c (Treating moderate AD)	Existing (i.e. prevalent case) AD patients with ***moderate AD*** (MMSE = 20 to 10)	Best supportive care (BSC) Donepezil Rivastigmine (×2)* Galantamine	As for Decision problem 1a.
Decision Problem 2a (Treating people with moderate and severe AD)	Existing AD patients with ***moderate to severe AD*** (MMSE = 20 to 0)	Best supportive care Memantine	Those who start on BSC, stay on it Drug treatment is continued until the clinical decision to stop treatment (e.g. no longer responding).
Decision Problem 2b (Treating people with severe AD)	Existing AD patients with ***severe AD*** (MMSE <10)	Best supportive care Memantine	Those who start on BSC, stay on it Drug treatment is continued until the clinical decision to stop treatment (e.g. no longer responding).
Decision Problem 3 (Treating people with moderate AD)	Existing AD patients with ***moderate AD*** (MMSE = 20 to 10)	Best supportive care Donepezil Rivastigmine (×2)* Galantamine & Memantine (if trial data for moderate only)	Those who start on BSC, stay on it Drug treatment is continued until either (a) clinical decision to stop treatment (e.g. no longer responding) or - for AChEIs - (b) patient progresses to severe AD.

* Rivastigmine patches and rivastigmine capsules were modelled as separate comparators due to the different mean daily costs of the two product types at typical doses.

Note that in Decision Problem 1 there is no option of switching to memantine when patients progress from mild to moderate disease, even though this is a possibility under the current licensed indications for memantine. The cost-effectiveness of each drug is assessed using the doses reported in the RCTs.

7.2. Overview of decision model development

The process of developing a decision model for this technology assessment had five main stages. These were:

1. Preparation and familiarisation

   (a) Familiarisation (by JP and RA) with past economic modelling studies in Alzheimer's disease, and in particular the perceived strengths and weaknesses of the modelling approaches used by the manufacturers and the technology assessment group in the 2004.

   (b) Rapid reviews to re-assess what factors drive (or what are associated with) changes in care costs or changes in health-related quality of life during the progression of Alzheimer's disease.

   (c) Contact with experts in the field.

2. Choosing between discrete event simulation or Markov (discrete state) modelling.

3. Exploring the possible development of a 'two-dimensional' Markov model of Alzheimer's disease progression – that is, a natural history disease model which simulated change through stages of both cognitive status and either functional status or behavioural symptoms. Given the typically univariate reporting of trial outcomes, this approach would probably require access to individual patient data.

4. Taking the 2004 SHTAC-AHEAD model from the previous technology assessment - which was based around a multivariate model for predicting time to full-time care - and both updating the model parameters, and adapting the model to try and address some of the more substantial criticisms made of it.

5. Developing a new Markov model, which is structurally quite similar to the SHTAC-AHEAD model, but has been based on a time-to-institutionalization equation based on a cohort of UK Alzheimer's patients.

7.2.1. Preparation and familiarisation

JP (decision modeller) and RA (health economist) obtained and read the previous technology assessment report, and related subsequent economic analyses. We also obtained various journal articles (pre- and post-2004) which discuss the area of modelling Alzheimer's disease.

In parallel with this, we (JP, HH, and RA) conducted several small reviews to get an evidence-based overview of the factors which are either associated with the Alzheimer's disease care cost or those factors which are associated with health-related quality of life (or utility) in Alzheimer's disease. For both reviews, while we were interested in what factors have been found to be associated with costs or quality of life in general, we were more specifically interested in which events or stages of disease progression appeared to be most associated with temporal changes in costs or quality of life. We were particularly interested in which clinical events, or main stages of disease progression, or changes in a patient's home circumstances etc. lead to step-change in health or social care costs. For example, the main model used in the previous technology assessment of these drugs for NICE in 2004, assumed that the major leap in health care costs was when patients required full-time care. The purpose of these reviews was to allow us to revisit such assumptions in the light of relevant published evidence.

The review questions which directed our searching and selection of studies are shown in *Table 104* below. The findings relating to published research on quality of life and utility in Alzheimer's disease are summarised in section *7.3.9*.

The review of published studies relating to the cost of care for Alzheimer's disease identified three recent key reports which estimated the cost of care in the UK (Dementia UK report in 2007; the National Audit Office's 2007 report into improving services for people with dementia, and; the more recent Dementia 2010 report including a rigorous and up-to-date cost-of-illness study).[5-7] However, other published studies relating to the cost of care for people with dementia or Alzheimer's disease in the UK mostly pre-dated the previous

technology assessment and have been summarised in it (Kavanagh and Knapp, 2002; Lowin et al. 2001; Souetre et al. 1999: Wolstenholme et al. 2002, and; Livingston et al. 2004).[178-182] Apart from closely reading two very recent systematic reviews of international evidence (mainly of cost-of-illness studies) relating to the cost of care for Alzheimer's disease,[183;184] we did not review in detail the other recent cost-related studies from the USA or other countries that we found.[185-187]

TABLE 104 *Questions which underpinned our background preparatory reviews*

Questions for background review on costs of AD	Questions for background review on quality of life in AD
▪ Which clinical events, or main stages of Alzheimer's disease progression - or changes in a patient's living situation - lead to a step-change in health or social care costs (especially in the UK)? ▪ Which markers or measures of Alzheimer's disease progression (e.g. cognitive function, functional ability, behavioural or psychotic symptoms, physical health), either individually or in combination, are most predictive of health and/or social care costs (especially in the UK)? ▪ In England and Wales, what are the typical stages or pathways of care for people with Alzheimer's disease? ▪ In England and Wales, to what extent are the costs of caring for people with Alzheimer's disease borne by (i) the NHS (ii) Personal Social Services (iii) local authorities (iv) other organisations such as voluntary organisations?	▪ Which clinical events or main stages of Alzheimer's disease progression - or changes in a patient's living situation – are associated with step-changes in people's health related quality of life or utility? ▪ Which markers or measures of Alzheimer's disease progression (e.g. cognitive function, functional ability, behavioural or psychotic symptoms, physical health), either individually or in combination, are most predictive of the quality of life (or and utility) of people with AD? ▪ Which markers or measures of Alzheimer's disease progression (e.g. cognitive function, functional ability, behavioural or psychotic symptoms, physical health), either individually or in combination, are most predictive of the quality of life (or and utility) of the carers of people with AD?

7.2.2. Discrete event simulation v. Markov modelling

On the basis of the reading and discussions as part of our preparation and familiarisation work, it was clear that an early decision had to be made about whether to model using a Markov (discrete disease state) modelling approach, or whether to use discrete event simulation methods. Although some analysts have argued that the complexity of Alzheimer's disease - particularly heterogeneity of disease progression between patients and competing health risks - is best reflected using a discrete event simulation approach,[153;188] we decided this was not likely to be suitable for our modelling task for the following main reasons:

- In general, the kind of information reported in published trials is difficult to use in discrete event simulation models – those who have used discrete event simulation methods tend to have access to individual patient data from trials.
- As an independent technology assessment group we very rarely have access to individual patient data from trials, and we had no reason to believe that obtaining such data for this technology assessment would be likely

We concluded that the potential delay to our model development (due to new software acquisition and training), combined with also building in a reliance on access to individual patient data, were incompatible with the time and other limitations of developing a model as part of the NICE technology assessment process. Therefore we decided that some form of multi-dimensional Markov model should be our initial choice of modelling approach.

7.2.3. Exploring the feasibility of a multi-dimensional Markov model of Alzheimer's disease progression

In 2004, but even more so by 2010, there is a clear view amongst most who have attempted to model Alzheimer's disease, or reviewed cost-effectiveness modelling in the disease area, that basing a disease progression model on decline in cognitive status alone is partial and inadequate.[189;190] More specifically, cognitive status alone is generally not highly associated with health-related quality of life or costs. Research evidence confirms the clinical view that Alzheimer's disease is a complex multi-dimensional disease, and therefore that any comprehensive model of disease progression in Alzheimer's disease should aim to capture changes in:

- Cognitive status
- Functional status (e.g. activities of daily living)
- Behavioural difficulties

We therefore decided to devote several weeks to exploring the feasibility of developing at least a two-dimensional model of Alzheimer's disease, while considering both the limitations and heterogeneity of trial outcome data available, and also the lack of standardised methods for deriving transition probabilities and treatment effect estimates from outcome data for two or more outcomes at the same time. (NB. We decided early that aiming for a three-

dimensional model would be unfeasibly ambitious given the timescales within which technology assessments for NICE have to be produced.) A fuller description of this modelling feasibility assessment is presented in Appendix 16. Below we summarise the main issues explored and our conclusions.

It was nevertheless clear that the first dimension of Alzheimer's disease for our model probably needed to be cognitive status. Cognitive status measured as either MMSE or ADAS-cog is both the most commonly reported outcome in the trial literature and is also the measure of disease severity on which marketing licences are based. It is also specified in the NICE scope as a possible basis for subgroup analysis.

We then reviewed the identified trial and other research literature to investigate whether either functional status or patient behaviour were valid and reliable independent predictors of quality of life or care costs. This led to us identifying behavioural status as the likely second dimension, although functional status was not totally ruled out and the diversity of different measures of behavioural status in the published trials remained a concern.

We then explored the following essential modelling requirements in more detail:

1. How to derive the transition probabilities for the best supportive care disease progression model from available data

2. How to obtain defensible estimates of treatment effect across two outcome dimensions.

3. How to obtain defensible estimates of health state utilities for both cognitive and functional or behavioural status.

The detail of our work exploring how to meet these requirements, and in particular what data would be needed, is described in Appendix 16. This included making a data request to the manufacturers for individual patient data (from the control arms of the main effectiveness RCTs) that was submitted to them via NICE. However, when in early March 2010 we were given acces to the individual patient data of two UK-based cohort studies of people with Alzheimer's disease (see below) this data request was retracted.

Ultimately we reached a time-point in the process where we had to assess whether these various modelling challenges (and especially their related data requirements) would be

solvable in time to allow the proper production and description of the cost-effectiveness results for NICE by the middle of June. We made the judgement in March that there was a high risk that we would not be able to adequately resolve these challenges, and therefore reverted to the development and improvement of the model developed for the 2004 technology assessment.

7.2.4. Making improvements to the SHTAC-AHEAD model

We first tried to replicate the SHTAC-AHEAD model and its cost-effectiveness results. We then began the process of making more step-by-step improvements to the SHTAC-AHEAD model, beginning with relatively straightforward changes (such as extending the time horizon, and updating the discount rates to those that currently apply to UK public sector cost-benefit analysis) and then making other changes to try to address other previously identified limitations of the SHTAC-AHEAD model.

The order in which we chose to tackle different potential improvements to the SHTAC-AHEAD model was guided by a complete list of previously identified limitations of the model, compiled from NICE documentation, research papers and a conversation with the health economist who worked of the previous model at SHTAC (Dr Colin Green). These limitations (see Appendix 17) were then prioritised according to (i) the perceived importance of the stated limitation in the context of producing a cost–utility analysis according to NICE methods guidance, and (ii) the expected simplicity/difficulty of developing a solution to the limitation. Using this 'master list' of identified limitations we also consulted NICE about which ones they thought would be the most important to try and address.

At the same time, the model was being constructed to be flexible; so that if individual patient data about UK Alzheimer's disease patients become available to us, we could then base the model around time-to-event data from this data, rather than the US data used to drive the time-to-FTC equations in the SHTAC-AHEAD model. (The reliance on the US data had been identified as one of the more important perceived limitations of the previous modelling.)

7.2.5. Building a time-to-institutionalization model based on UK data

Following correspondence with the authors and principal investigators, on the 5th of March 2010 we were kindly sent the full data set of the London-based LASER-Alzheimer's disease

study (Principal Investigator, Prof Gill Livingston, UCL, London), and on the 9th of March we were kindly sent the full data set of the Oxfordshire Alzheimer's disease data (health economist Dr Jane Wolstenholme, University of Oxford).

The availability of both these UK datasets of individual patient data about Alzheimer's patients and their care and outcomes opened up a number of possibilities for our modelling. In particular, it provided the possibility of using UK data to develop a multivariate regression model of time-to-institutionalisation (or time to full time care) to replace the US (AHEAD) study-based equations in the SHTAC-AHEAD model. Importantly also, it allowed us to explore for ourselves possible relationships between time-to-institutionalisation and MMSE, and care costs, with a view to further informing model assumptions about gradually increasing care costs, and gradually decreasing health-related quality of life in the time before patients become institutionalised. Again, a key criticism of the previous economic model was that QALY gains were only achieved for patients who survived to entering the full-time care state. Similarly though, the previous model did not allow the possibility that in the years and months leading up to the point of needing full-time care, costs of care would be likely to increase over time with disease progression. As a health technology assessment team, we were very keen to find a way of addressing these two weaknesses of the previous modelling approach.

The following sections describe the methods, and then results, of the final modelling approach that we developed. Ultimately, for various reasons, we made more use of the Oxfordshire Alzheimer's disease dataset than the LASER-AD study data, and we explain how and why in the relevant sections.

7.3. Methods

7.3.1. Model structure

The above development stages therefore led to the development of a decision model based broadly on the structure of the three-state Markov model described in the previous TAR[2]. An exploration of the SHTAC model in light of the various criticisms and issues raised during and since the 2004 review process, has led to the development of a model based upon time to institutionalization, parameterised with updated estimates of effectiveness, costs and utilities. A review of all documentation (from manufacturers, interest groups, NICE and the

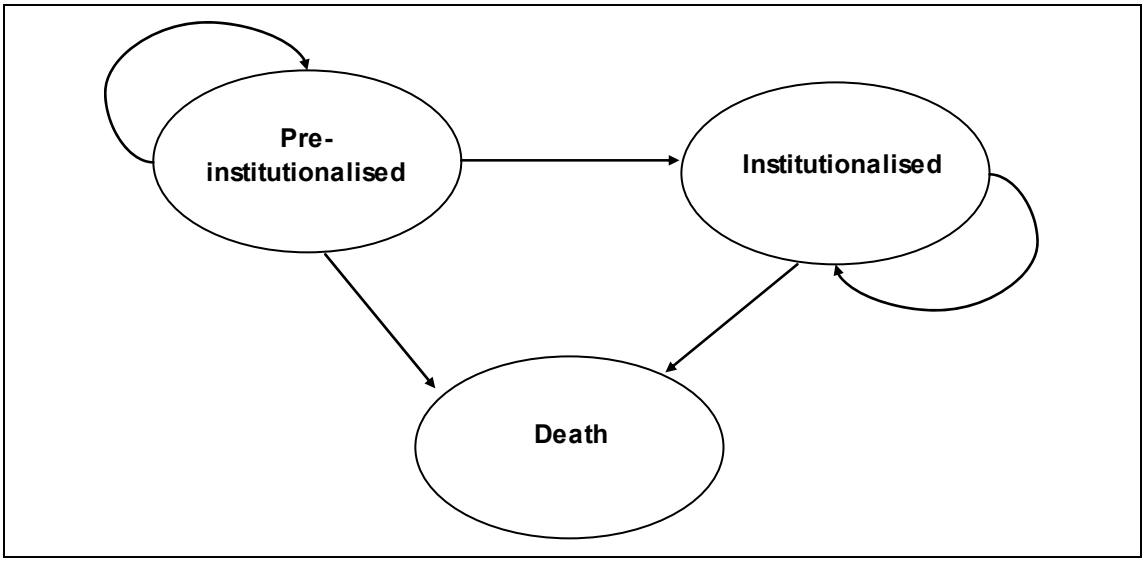
published literature) relating to the decision model described in the previous TAR was undertaken. From this review a list of the various criticisms and issues associated with the SHTAC model was created and is shown in Appendix 17. Using this list a number of changes to the SHTAC model structure and the parameter values used in the three-state Markov model were explored. These alterations are described in the appropriate sections below. The model was developed in Microsoft Excel 2007 with additional analyses undertaken in the statistical software package R.

7.3.2. Model states

The Markov model consists of three states: pre-institutionalization, institutionalization and death (see *Figure 59*). Note that this model differs from that of the SHTAC model since progression is based on time to institutionalization in the PenTAG model, not time to full-time care (defined by "equivalent institutional care"[191] including day and night "supervision of personal care, safety or medical care"[192]) as it was in the SHTAC model. Institutionalization is defined in the IPD from Wolstenholme and colleagues, and thus in the PenTAG model, as "Living in a residential home or a nursing home (not as short respite care) or in hospital on a long-term or permanent basis" (personal communication with Jane Wolstenholme). A particular criticism of the SHTAC model was whether the risk equations used to predict full time care from the US study by Stern and colleagues[191] could be generalised to England and Wales. Furthermore, the definition of full time care was queried re its relationship to institutionalization, the major step change in costs associated with Alzheimer's disease as identified by a review of costs in Alzheimer's disease mentioned above. In an attempt to address these concerns, the PenTAG model is based upon UK data predicting time to institutionalization. Depending on severity, at the beginning of the model individuals in the cohort start in either the pre-institutionalization state or the institutionalization state, e.g. for the base-case analysis for mild to moderate severity (decision problem 1a, see *Table 103*) 90% of the cohort are assumed to start in the pre-institutionalization state. Transition to death from either of the alive states can occur at any point in time. It is assumed that once an individual becomes institutionalized they do not return to the pre-institutionalized state, thus there are no backward transitions in this model. Note that due to treatment discontinuations some individuals may be on treatment while in the pre-institutionalized state while others may not be on treatment (this is further explained in Section *7.3.7.2*). The three state Markov model was applied to a cohort of 1000 individuals with mild to moderate

Alzheimer's disease to model the cost–utility of the AChEIs (decision problem 1a in *Table 103*) and moderate to severe Alzheimer's disease to model the cost–utility of treatment with memantine (decision problem 2a in *Table 103*). Information on the characteristics of the modelled population is given in the section below.

FIGURE 59 *Diagram of the three-state Markov model*



7.3.3. Modelled population

For the three cholinesterase inhibitors, donepezil, rivastigmine (capsules and patches) and galantamine, the base-case analysis modelled a cohort of people with mild to moderate Alzheimer's disease (MMSE 26-10). For memantine, the base-case analysis concerned people with moderate to severe Alzheimer's disease (MMSE 10-20). In exploratory sensitivity analyses, the cost-effectiveness of treatment with donepezil, rivastigmine (capsules and patches) and galantamine was investigated for a cohort of people with mild Alzheimer's disease. Further exploratory sensitivity analyses investigated the cost–utility of donepezil, rivastigmine (capsules and patches), galantamine and memantine for people with moderate only Alzheimer's disease (MMSE 10-20) and the cost-effectiveness of memantine in the treatment of people with severe only Alzheimer's disease (MMSE <10).

A prevalent cohort is assumed for this decision model since the data informing disease progression for untreated patients are from a prevalent cohort. Thus, the decision problem only considers the costs and QALYs of a treatment change for the prevalent cohort of individuals with Alzheimer's disease. It does not consider the costs and QALYs of individuals

diagnosed in the future with Alzheimer's disease, which will typically differ from patients in the prevalent cohort[193]. It is therefore necessary to state from the outset that this model is based on an assumption that individuals have had a diagnosis of Alzheimer's disease for a mean of 4.9 years (IPD from Wolstenholme and colleagues[181]). Data regarding the characteristics of people with Alzheimer's disease were primarily based on IPD from the study by Wolstenholme and colleagues[181]. This study is used to inform much of the PenTAG decision model (including disease progression and cost estimates; see Sections *7.3.10*, *7.3.9* and*7.3.10*). It was chosen as it contains data on untreated people with Alzheimer's disease in England and was made available to us by Wolstenholme and colleagues [181]. A UK-based epidemiological cohort study, such as that by Wolstenholme and colleagues was preferred over clinical trial data to avoid any biases of assuming disease progression based on RCT populations which are subject to a number of inclusion and exclusion criteria not representative of our target population: people with Alzheimer's disease in England and Wales. Furthermore, longer follow-up data was available from the Wolstenholme dataset than that available from clinical trial data. A second UK-based epidemiological dataset was available from the LASER-AD study [182]. This study was not used to predict disease progression as many participants were taking cholinesterase inhibitors and/or memantine during the study period. However, data from the LASER-AD study was used to justify and/or corroborate a number of assumptions in the model.

The 1997/8 UK-based study by Wolstenholme and colleagues[181] provided estimates of the NHS and PSS costs associated with Alzheimer's disease. This was a retrospective cohort analysis of people diagnosed with Alzheimer's disease or vascular dementia. Having access to the IPD from this dataset made it possible to restrict all analyses to only those people with Alzheimer's disease (excluding eight out of 100 individuals who had vascular dementia). The study participants were recruited through GPs, community psychiatric nurses and consultant geriatricians in the Oxfordshire area during 1988-9. Up to 11 years follow-up data is available from this cohort. This data represents a prevalent cohort of 92 patients with Alzheimer's disease. At the time of study entry, patients were diagnosed with Alzheimer's disease a median of 4.0 years and a mean of 4.9 years ago.

For each patient, the time from study entry to institutionalization and death was recorded; 82 of the 92 patients died before the end of the study; 16 patients died before becoming institutionalized and 72 of the 92 patients were institutionalized. At the time of study entry,

among a number of outcome measures, the MMSE, Barthel ADL index and age of the patient were recorded.

The population characteristics from an analysis of the IPD from Wolstenholme and colleagues are shown in *Table 105*. These values were used in the PenTAG model to inform various parameter values for the base-case analyses. In exploratory sensitivity analyses, the cost-effectiveness of treatment with donepezil, galantamine and rivastigmine (capsules and patches) in a cohort of people with characteristics of mild Alzheimer's disease was assessed (decision problem 1b in *Table 103*). As was the cost-effectiveness of memantine compared to best supportive care for a cohort of people with severe Alzheimer's disease (decision problem 2b in *Table 103*) and the cost-effectiveness of donepezil, galantamine, rivastigmine (capsules and patches) and memantine in a cohort of individuals with the characteristics of moderate Alzheimer's disease (decision problem 3 in *Table 103*). The parameter values for the population characteristics from the Wolstenholme IPD used in the base-case and explorative sensitivity analyses are shown in *Table 105*.

TABLE 105 *Baseline population characteristics from a re-analysis of Wolstenholme and colleagues*[181]

	Severity of Alzheimer's disease				
	Mild to moderate (MMSE 26–10)	**Moderate to severe (MMSE 20–0)**	**Mild (MMSE 26–21)**	**Moderate (MMSE 20–10)**	**Severe (MMSE 9–0)**
N	71	70	22	49	21
Mean age	77.7	78.57	76.55	78.22	79.38
Mean MMSE	17	11.73	23.04	14.43	5.43
Mean Barthel ADL	17.52	16.34	18.88	16.94	14.92

Note that the data informing disease progression are that from a prevalent cohort of patients living in the community[181], and is therefore not fully representative of the target population of patients in England and Wales living in the community and in institutionalized care. It was therefore felt that the model should account for the fact that some individuals in the prevalent cohort are likely to be in institutional care. Data indicating the proportion of people with Alzheimer's disease who are institutionalized was available from the LASER-AD study. Livingston and colleagues[194] reported that 5.6% of individuals with MMSE >19, 27.1% of individuals with MMSE 15-19 and 59% of individuals with MMSE<15 were in institutional care at baseline. This translates to 13% for MMSE >14 (slightly different to the usual definition of MMSE>9 for mild to moderate disease) and 46% for MMSE < 20 (the usual definition for

moderate to severe disease). However, it is unclear from the LASER-AD study whether a prevalent or incident cohort are described and analysed, therefore there are questions as to how the baseline characteristics of the LASER-AD population compare to the baseline characteristics of the Wolstenholme study. In addition to this, recent evidence indicates that the number of individuals in institutional care is falling (see Knapp and colleagues[6], Chapter 4, page 50). Therefore in one-way sensitivity analyses, the LASER-AD results are used as a guide to assume that 10% of the mild to moderate cohort and 40% of the moderate to severe cohort are institutionalised at the start of the model. In the PenTAG model, both time to institutionalization and death are significantly dependent on age (see Section *7.3.8*), and so the cohort model allows three subgroups defined by age to be included, allowing some degree of heterogeneity to be modelled within the cohort. *Table 106* shows the baseline population characteristic parameter values for mild to moderate and moderate to severe cohorts. Note that, as would be expected, the more severe cohort has a slightly older profile. The parameter values are assessed for their impact on the cost–utility findings in one-way sensitivity.

TABLE 106 *Baseline parameter values for population characteristics in the base-case analyses*

	Mild to moderate Alzheimer's disease	**Moderate to severe Alzheimer's disease**	**Source**
Mean age			
Age group 1	69	69	Wolstenholme et al[181] IPD
Age group 2	77	78	
Age group 3	86	87	
Proportion in age group			
Age group 1	0.25	0.25	Wolstenholme et al[181] IPD
Age group 2	0.50	0.50	
Age group 3	0.25	0.25	
Mean MMSE	17	11.73	Wolstenholme et al[181] IPD
Mean Barthel-ADL	17.52	16.34	
Proportion starting in institutionalization	0.1	0.4	Informed by data from LASER-AD study [194]

7.3.3.1. Model population parameters and assumptions used in sensitivity analyses

In one-way sensitivity analyses, the characteristics of the modelled cohort for mild to moderate and moderate to severe analyses were changed to represent a different cohort using baseline data from the LASER-AD study. The alternative mean parameter values for the age, MMSE and ADCS-ADL of the cohort are shown in *Table 107*. Participants in the LASER-AD study are older than those from the Wolstenholme study, but the mild to moderate cohort from the LASER-AD study has slightly less cognitive impairment than that from the Wolstenholme study (MMSE: 19.21 from LASER-AD versus 17.52 from Wolstenholme study), while the moderate to severe cohort are slightly more cognitively compared in the LASER-AD study compared to the Wolstenholme study (MMSE: 10.91 from LASER-AD versus 11.73 from Wolstenholme study).

TABLE 107 *Parameter values used in sensitivity analyses from the LASER-AD study*

	Severity of Alzheimer's disease	
	Mild to moderate (MMSE 26–10)	**Moderate to severe (MMSE 20–0)**
Mean age		
Age group 1; 25% cohort	71	72
Age group 2: 50% cohort	81	82
Age group 3: 25% cohort	90	91
Mean MMSE	19.21	10.91
Mean ADCS-ADL	44.52	27.59

As noted above, further sensitivity analyses on the modelled cohort were undertaken to explore decision problems 1b, 1c, 2b and 3 as defined in *Table 103*. The parameter values for the cohort of mild, moderate and severe Alzheimer's disease patients are shown in the final three columns of *Table 105*. However, caution is needed when interpreting the results from these exploratory analyses, as the populations from which the estimates of effectiveness were obtained were not restricted to the mild, moderate to severe populations. For instance, only one RCT reports effectiveness data within a mild population[109], and this is only for one drug (donepezil) and for only one outcome (MMSE). Since the estimate for an effect on MMSE reported in Seltzer and colleagues[109] is similar to the pooled estimate for the mild to moderate cohort but with greater uncertainty (compare 1.25 (95%CI: 0.17, 2.33) from Seltzer and colleagues[109] with 1.24 (0.81, 1.66) in *Table 108*), an exploration of the cost–

utility of donepezil, galantamine and rivastigmine in a mild population was undertaken using the same effectiveness estimates as that for the mild to moderate population, only the characteristics of the population are altered. Therefore, the sensitivity analysis results for a mild cohort should be interpreted with caution. Similarly, the comparisons between drugs in a moderate cohort should be interpreted with caution since the populations in the RCTs informing the effectiveness are not entirely comparable, those reporting effect estimates for the AChEIs are based on mild to moderate populations, while that for memantine is from a single study with a moderate population.

7.3.4. Model assumptions

The model starts when treatment begins for the treated cohorts (point A in *Figure 60*). For the initial treatment period (point A to B), mean time to institutionalization and mean time to death are predicted using mean baseline characteristics of the cohort. After the initial treatment period (point B), any treatment effects are assumed to have occurred, and so from point B onwards, mean time to institutionalization (point C) is predicted based on the mean baseline characteristics plus the mean treatment effect for the treated cohorts. For example, if a mean baseline MMSE of 17 and a mean treatment effect of 0.5 on the MMSE scale are assumed, the mean time to institutionalization for an untreated cohort would be predicted using a mean MMSE of 17. Mean time to institutionalization for a treated cohort is based on a mean MMSE of 17 for the initial treatment period, but would be based on a mean MMSE of 17.5 from point B onwards. This leads to treated cohorts having a delay in institutionalization compared to best supportive care. The length of this treatment period (point A to B) depends on the length of follow-up reported in the source RCTs. Note that some patients continue to be treated after point B.

In the 2004 SHTAC model, only costs during the initial treatment period were accrued, not utilities (see number 27 in Appendix 17). In the PenTAG model both treatment costs and utilities are accrued during the initial treatment period.

FIGURE 60 *Time-line of model for typical individual with Alzheimer's disease*

A: Model starts; B: Prediction of time to institutionalization incorporating treatment effect on MMSE and Barthel-ADL; C: Individual becomes institutionalized.

Since time to death is predicted by age, MMSE and ADL, and given that the treatments affect MMSE and ADL, it is possible to assume that the treatments delay death as well as delaying institutionalization. There is, however, no evidence from the RCTs that treatment increases survival. Neither is there any epidemiological evidence to suggest a treatment effect on survival. Therefore, for the base-case analysis, it is assumed that treatment with donepezil, rivastigmine (capsules and patches), galantamine or memantine delays time to institutionalization, but has no impact on survival. In sensitivity analyses, a treatment effect on survival is assumed and the results presented in similar detail to the base-case analyses.

A criticism of the SHTAC model from the previous MTA (see number 9 in Appendix 17) was that there was no daily benefit accrued by treated individuals prior to the point of needing full-time care. In the SHTAC model any individuals dying while in the pre-full time care state or remaining in the pre-full-time care state at the end of the model did not contribute any health benefit, yet accrued treatment costs. To overcome this criticism, IPD from the Oxfordshire study by Wolstenholme and colleagues was used to refine the pre-institutionalized state of the PenTAG model to allow for gradual increases in costs and gradual reductions in health-related quality of life during the pre-institutionalized state. Therefore, rather than a single utility or cost value being assigned to the pre-institutionalization state (as in the SHTAC model), pre-institutionalized utility and cost are dependent upon time to institutionalization. This allows individuals predicted to be close to being institutionalized to have higher costs and lower utility compared to those individuals who are predicted to be years away from being institutionalized (see Sections *7.3.10* and *7.3.9* for further detail).

The PenTAG model allows for treatment discontinuations (see Section *7.3.7.2*), and assumes that for the three cholinesterase inhibitors, treatment stops once they enter

institutionalization. Thus, the model implicitly assumes that institutionalization is equivalent to severe Alzheimer's disease (MMSE < 10). Therefore, once in an institution, patients' quality of life and utility are assumed to be that of people with severe Alzheimer's disease (MMSE <10). At MMSE<10 both the marketing licence and current guidance recommends that patients be taken off the cholinesterase inhibitors. This equivalent assumption was made in the SHTAC model for patients entering full-time care and criticised (see number 2 in Appendix 17), however analysis of the IPD from Wolstenholme and colleagues suggests that entering institutionalization is a good proxy for severe Alzheimer's disease (as measured by the MMSE): the mean time at which participants reached MMSE of 9 is 0.04 years prior to institutionalization. No such assumption is required to model memantine, as the drug is licensed for moderate to severe Alzheimer's disease, therefore unless treatment is discontinued (see Section *7.3.7.2*), memantine is assumed to be taken by individuals until they die.

7.3.5. Time horizon

A monthly time cycle was used in the model and the time horizon was 20 years. By this time it was estimated that <5% of the cohort would be alive.

7.3.6. Discount rates

In base-case analyses discount rates of 3.5% were applied to both costs and health benefits. In sensitivity analyses differential discount rates were explored such that health benefits were discounted at 1.5% and costs at 3.5%.

7.3.7. Sources of effectiveness data

7.3.7.1. Clinical effectiveness

For estimates of clinical effectiveness the highest quality evidence was required. Therefore only estimates from those RCTs identified in Section 5 contributed to the parameterisation of the model. The estimates of clinical effectiveness sought were those from head-to-head trials (as per NICE methods guide[195]), reporting on

- cognition, MMSE in particular, and

- functional ability, ADCS-ADL in particular.

The longest follow-up consistent across the different drugs and outcomes was six months. Therefore treatment effect estimates at this time-point were used in the base-case analysis, and so the time between points A and B in *Figure 60* is six months. Longer follow-up data for donepezil were available and were assessed in sensitivity analyses when compared directly to best supportive care. The longer follow-up data were not used to compare across other AChEIs since the effectiveness data would not be comparable.

Estimates of effectiveness as measured on the MMSE and ADCS-ADL scales, reported in Section 4 were used in the cost-effectiveness modelling. For treated cohorts, the mean difference in MMSE and ADCS-ADL from RCTs was applied to the baseline estimate of MMSE and ADCS-ADL used in the best supportive care cohort. Thus, the mean baseline MMSE score for treated cohorts was expected to be greater than that for the best supportive care cohort since a larger MMSE indicates better cognitive function than a smaller MMSE score. Similarly, the mean difference in ADCS-ADL from the RCT evidence was added to the mean ADCS-ADL score in the best supportive care cohort, with ADCS-ADL scores expected to be higher in treated cohorts. The effectiveness estimates used in the decision model are given in *Table 108*. Note that the only RCT providing effectiveness evidence on ADAS-cog for memantine did not restrict participants to use of memantine only, participants also received AChEIs (Section 4.8), therefore these data are not included *Table 108*.

TABLE 108 *Estimates of effectiveness (at 6 months) used in the PenTAG decision model*

	Outcome measure	WMD (95% CI)	Analysis type	Source
Donepezil (10mg)	MMSE	1.24 (0.81, 1.66)	M-A result	AD2000 (2004)[96], Rogers et al (1998)[106], Gauthier et al (2002)[98], Seltzer et al (2004)[109], Mohs et al (2001)[103], Winblad et al (2001)[110] (Appendix 5, Figure 15)
	ADCS-ADL	2.02 (1.06, 3.28)		Average of estimate from galantamine (24mg) and rivastigmine (≤12mg)
	ADAS-cog	-2.90 (-3.61, -2.18)	M-A results	
Galantamine (16-24mg)	MMSE	1.13 (0.72, 1.54)		Average of donepezil (10mg) and rivastigmine (≤12mg)
	ADCS-ADL	2.23 (1.33, 3.14)	M-A result	Tariot et al (2000)[119], Brodaty et al (2005)[89] (Figure 26)
	ADAS-cog	-3.05 (-3.52, -2.57)	M-A result	
Rivastigmine capsules (9-12mg)	MMSE	1.02 (0.63, 1.41)	M-A result	Feldman & Lane (2007)[132], Winblad et al (2007)[134] (Figure 35)
	ADCS-ADL	1.80 (0.20, 3.40)	Single study	Winblad et al (2007)[134]

	ADAS-cog	-2.34 (-3.38, -1.30)	M-A result	
Rivastigmine patches (10cm^2)	MMSE	1.10 (0.52, 1.68)	Single study	Winblad et al (2007)[134]
	ADCS-ADL	2.20 (0.62, 3.78)	Single study	Winblad et al (2007)[134]
	ADAS-cog	-1.60 (-2.73, -0.47)	Single study	Winblad et al (2007)[134]
Memantine (15-20mg)	MMSE	0.70 (0.02, 1.38)	Single study	Reisberg et al (2003)[136]. Note: only data from memantine v placebo RCTs
	ADCS-ADL	1.41 (0.04, 2.78)	M-A result	Reisberg et al (2003)[136], Van Dyck et al (2007)[137] (Figure 46) Note: only data from memantine v placebo RCTs

The sources of effectiveness were also decided upon by comparison with dose levels reported in RCTs and those available in the BNF. For instance, there was effectiveness evidence for 32mg/day of galantamine, however this was not available as a dose regime in the BNF. Similarly, although a daily dose of 36mg/day of galantamine was available, this was not recommended by the BNF. Therefore effectiveness data regarding galantamine at these dose levels (32mg/day and 36mg/day) were excluded from being considered as inputs for clinical effectiveness in the cost–utility model.

To help obtain informative findings for the different drugs, it was felt that an assessment of the cost–utility would be more appropriate for defined dose levels rather than considering a large mix of doses. Dose levels used in the RCTs providing MMSE and ADCS-ADL outcomes for each drug were noted.

- Donepezil: daily doses were reported to be 5mg or 10mg in the RCTs described in *Table 108*, however the majority of RCTs reported treatment effects for participants receiving 10mg. Since donepezil at a daily dose of 10mg is included in the BNF 10mg was taken to be the dose of donepezil to be considered in the cost–utility analysis. The limited effectiveness evidence for 5mg dose is assessed in sensitivity analyses.

- Galantamine: two RCTs reported on the clinical effectiveness of galantamine as measured on the ADCS-ADL[89;119]. Both of these RCTs contained individuals taking 8-24 mg/day. In Brodaty and colleagues[89], treated participants received 8mg/day for the first 4 weeks, with treatment then titrated to a maximum of 24mg/day. The average daily dose received by participants and reported in Brodaty and colleagues was approximately 17mg. Since the BNF indicates that 16mg/day is the lower recommended dose the assessment of galantamine is assumed to be for 16-24mg/day

doses and the drug costs are calculated based on an average of the two doses (see Section *7.3.10.1* for further details on drug costs).

- Rivastigmine: two RCTs reported clinical effectiveness on MMSE or ADCS-ADL[132;134]. Feldman & Lane[132] only reported on the use of capsules (2-12mg/day), while Winblad and colleagues[134] reported on the effectiveness of both capsules (3-12mg) and patches ($10cm^2$ and $20cm^2$). For the capsules, in both Feldman & Lane[132] and Winblad and colleagues[134] participants were titrated from a dose of 2mg/day and 3mg/day, respectively to 12mg/day. The mean daily doses received by participants were reported to be approximately 9mg in Feldman & Lane, and 9.7mg in Winblad and colleagues. Therefore the drug costs are based on a combination of 9 and 12mg/day doses. Assessment of the mean differences between the $10cm^2$ and $20cm^2$ patch for MMSE and ADCS-ADL outcomes at 6 months suggested little difference between the effectiveness of the patches compared to placebo (see *Table 109*). Furthermore, the $20cm^2$ patch is not a dose regime in the BNF. Therefore all assessments of the rivastigmine patch are based on the $10cm^2$ patch. Comparison of the effectiveness and the costs of rivastigmine capsules and patches indicate that these are different technologies and so both are considered in the cost–utility analyses.

TABLE 109 *Mean differences in MMSE and ADCS-ADL scores for rivastigmine patches reported by Winblad and colleagues[134]*

	MMSE	**ADCS-ADL**
$10cm^2$ patch	1.1 (0.52, 1.68)	2.2 (0.612, 3.78)
$20cm^2$ patch	0.9 (0.32, 1.48)	2.3 (0.53, 4.07)

- Memantine: the two RCTs contributing to the effectiveness data for memantine (Reisberg and colleagues (2003)[136] and Van Dyck and colleagues (2007)[137]) both compared memantine to placebo without the additional use of cholinesterase inhibitors. The dose used in the RCT of Reisberg and colleagues is reported to be 20mg while in the RCT of Van Dyck and colleagues an initial dose of 5mg is assumed with subsequent incremental doses leading to the target dose of 20mg. Therefore, the cost–utility of memantine reported here is based on an average of the 15 and 20mg/day costs.

A consequence of using the UK dataset from Wolstenholme and colleagues is that functional capacity is measured on the Barthel ADL index, an index not used or reported in any of the included RCTs. To incorporate this information the effectiveness evidence from the ADCS-ADL scale used in the RCTs had to be translated onto the Barthel ADL index. The Barthel Index includes the following ADLs: toileting, bathing, grooming, dressing, feeding, transferring from a sitting to a standing position, mobility and use of stairs. We are not aware of a mapping from the ADCS ADL index to the Barthel index in the literature. It is tempting to assume a direct proportionality between scores on the ADCS ADL and Barthel indices, with the constant of proportionality equal to the ratio of the maximum score on the Barthel index (20) and the maximum score on the ADCS ADL index (78). However, this would be a strong assumption, and not evidence-based. Instead, we estimated a quadratic mapping from the ADCS ADL index to the Barthel index, using data from Galasko and colleagues (2005)[196], which gives the mean scores over 145 patients in the USA at each of 3 time points (t = 0, 6, 12 months) for each of the 19 questions of the ADCS-ADL *severe* index. From this data, we first estimated the corresponding scores at each of the three time points on the Barthel index. Next, we estimated the corresponding scores at each of the three time points on the ADCS-ADL index. Together this gave us three data points for the mapping from the ADCS-ADL to the Barthel index.

All data was taken from Table 1 in Galasko and colleagues (2005)[196]. First, the three Barthel scores, corresponding to times 0, 6 and 12 months were derived as follows. Each question of the Barthel index (bowels, bladder, grooming, toilet, feeding, transfer, mobility, dressing, stairs and bathing) was taken in turn. For each question, one question of the ADCS-ADL *severe* index was identified which most closely correlated with the question on the Barthel index. In many cases, there is an exactly analogous question on the ADCS-ADL *severe* index, e.g. bathing. Next, a simple relationship was derived between the score on the question on the ADCS-ADL *severe* and the score of the question on the Barthel index. This was usually, but not always, a simple direct proportional relationship, with the constant of proportionality equal to the ratio of the maximum score on the Barthel index and the maximum score on the ADCS-ADL *severe* index. For example, the ratio for bathing was set to 1/3, given a maximum score of 1 on the Barthel index and 3 on the ADCS-ADL *severe* index.

Next, the score for each time point for a given question on the Barthel index was calculated as the score for the correlating question on the ADCS-ADL *severe* index multiplied by the

relationship between the scores on the Barthel and ADCS-ADL *severe* indices described in the previous paragraph.

When no single question on the ADCS-ADL *severe* index could be identified which closely correlated with a given question on the Barthel index, we set the score for that question on the Barthel index equal to the maximum score for that question on the Barthel index multiplied by the ratio of the total score over all questions on the ADCS-ADL *severe* index and the maximum possible total score on the ADCS-ADL *severe* index (equal to 54). In this way, the scores on each of the ADCS-ADL *severe* questions influenced the score for the single question on the Barthel index. This procedure yielded the following values on the Barthel index, corresponding to patient times 0, 6 and 12 months respectively: 13.16, 10.33 and 7.74.

Next, the three ADCS-ADL scores, corresponding to times 0, 6 and 12 months were derived in the way, as described above for the Barthel scores, except when no single question on the ADCS-ADL *severe* index could be identified which closely correlated with a given question on the ADCS-ADL index, we set the score for that question on the ADCS-ADL index exactly as described above (i.e. equal to the maximum score for that question on the ADCS-ADL index multiplied by the ratio of the total score over all questions on the ADCS-ADL *severe* index and the maximum possible total score on the ADCS-ADL *severe* index (equal to 54)), *then multiplied by 50%.* It was necessary to multiply by 50% because only 50% of patients in the study of Galasko and colleagues (2005)[196] even attempted to answer the question. These questions on the ADCS-ADL index, for which there is no correlating question on the ADCS-ADL *severe* index (e.g. making a meal, shopping, reading, writing) are activities very rarely attempted by people with moderate to severe Alzheimer's disease. Indeed, these questions are omitted from the ADCS-ADL *severe* index because the ADCS-ADL *severe* index is designed for moderate to severe Alzheimer's disease. We acknowledge that the 50% factor is an approximation, but this is our best estimate given the lack of further data.

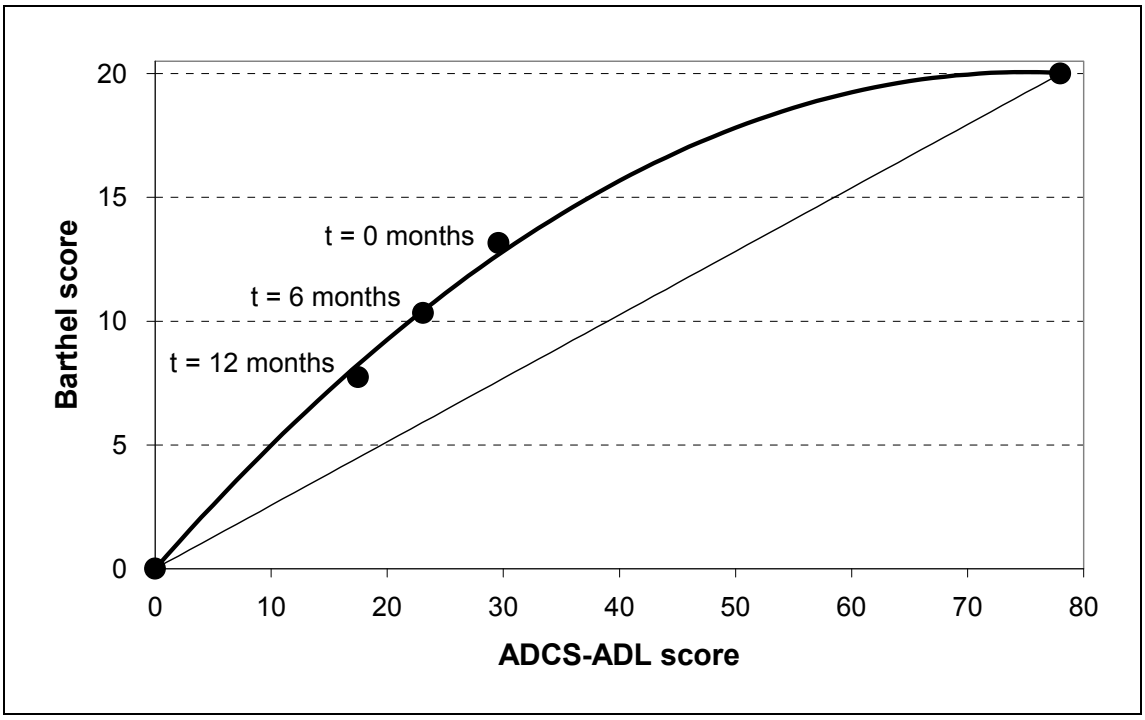
This procedure yielded the following values on the ADCS-ADL index, corresponding to times 0, 6 and 12 months respectively: 29.60, 23.09 and 17.48. In addition, we know that when the maximum score of 78 is achieved on the ADCS-ADL index, the maximum score of 20 must be achieved on the Barthel index.

We then fitted a statistical model in "R" software to these four data points with the Barthel score as the response variable and a quadratic in the ADCS-ADL score as the explanatory

variables, where the intercept was constrained to equal zero (as the function must pass through the origin). The deterministic mapping is given as follows (*Figure 61*);

*Barthel score = 0.534 * (ADCS-ADL score) - 0.0036 * (ADCS-ADL score)^2.*

FIGURE 61 *Statistical relationship between ADCS-ADL and Barthel ADL*



The thick curved line shows the relationship used in our base-case calculated from Galasko and colleagues.

For the PSA, it is necessary to model the uncertainty in the covariate coefficients. This was achieved as follows. First, the Cholesky matrix **C**, corresponding to the variance/covariance matrix of the parameter coefficients was calculated as

$$C = \begin{pmatrix} 0.0179 & 0 \\ -0.00024 & 0.00008 \end{pmatrix}$$

The rows and columns of *C* correspond to the linear and quadratic terms in the ADCS-ADL score respectively. Probabilistic covariate coefficients were then simulated as **y** + **Cz**, where **y** is the vector of coefficient means (given in the deterministic equations above), and **z** is a vector of independent standard normal variables[197].

There were two instances of missing data across the five treatments and two outcomes:

- An estimate of effect on ADCS-ADL at six months for donepezil (10mg), and
- An estimate of effect on MMSE at six months for galantamine (16-24mg).

It was assumed that this was a lack of evidence for an effect rather than a lack of effect. The average treatment effect from the same class of drugs was used for these two instances of missing data. In other words, the effectiveness of donepezil based on ADCS-ADL was taken as an average of the effectiveness of galantamine and rivastigmine capsules. Similarly, the effectiveness of galantamine based on MMSE was assumed to be an average of the MMSE effectiveness estimate from rivastigmine (capsules) and donepezil were assumed.

A treatment effect on cognition as measured by MMSE was used in the base-case as it was the scale used in the study by Wolstenholme and colleagues, and therefore used in the PenTAG model to predict time to institutionalization and time to death. In sensitivity analyses, a treatment effect on cognition as measured by the ADAS-cog score was assumed for all AChEIs and translated onto the MMSE scale using an equation published by Doraiswamy and colleagues[198]. Further sensitivity analyses explore the impact of increasing and decreasing the ADCS-ADL and MMSE effectiveness estimates for all technologies.

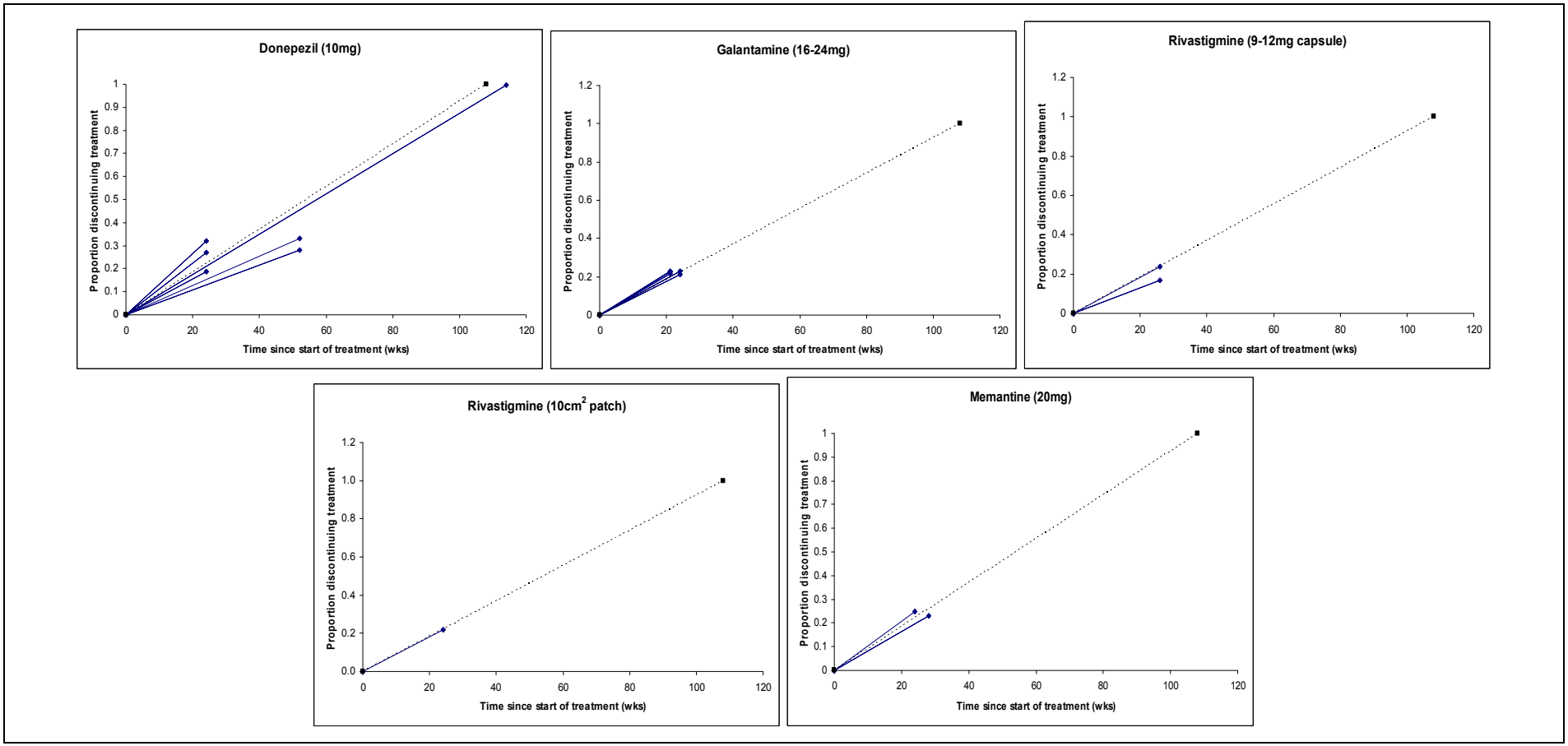
7.3.7.2. Treatment discontinuation

For all effectiveness estimates it is assumed that an ITT analysis has been undertaken, so that estimates relate to all participants and not only those continuing to take treatment. Given that many RCTs did not report an ITT analysis, this assumption is likely to over-estimate any treatment effects in the decision model.

Data on the proportions of individuals discontinuing treatment were available from the RCTs included in the systematic review. There was a great deal of information across different dose levels and follow-up, however each RCT only reported discontinuations at the last follow-up within each study. The available data are given in *Figure 62* for each drug. As can be seen the data are not entirely consistent across studies, with higher discontinuations observed at shorter follow-up than at longer follow-up (e.g. galantamine 16-24mg). In the base-case analysis a constant rate of treatment discontinuation was therefore assumed for all drugs at all doses. The basis for this value was a mixture of the evidence from the RCTs (indicating that if discontinuations carried on as reported at six months, by about two years most patients would have discontinued treatment) and clinical opinion on the length of time

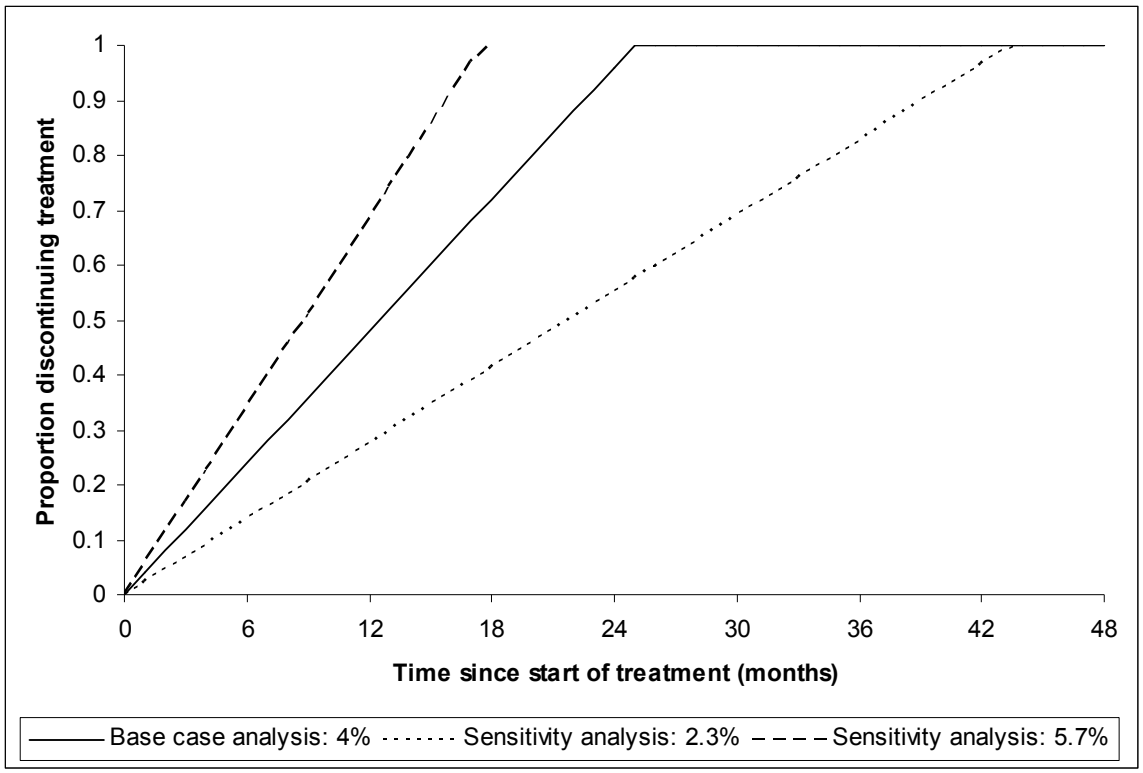
patients would generally spend on treatment. It is possible that the pattern of treatment discontinuations is not linear as assumed in the PenTAG model and that it may be more likely that many more patients discontinue treatment at the earlier stages of treatment. However, as noted, the only data available are that at a single time-point and a more complex relationship between discontinuations and time (other than a linear relationship) would require further, currently un-testable, assumptions to be made. It is therefore assumed that 4% of the total cohort discontinue treatment each month, so that after 2 years of treatment almost all individuals are no longer receiving treatment (see *Figure 63*). In the PenTAG model the proportion of participants discontinuing treatment was applied to the treatment and monitoring cost estimates. The impact of different discontinuation rates was assessed in sensitivity analyses. These were based on the minimum (2.3%) and maximum (5.7%) slopes across all technologies in *Figure 63*, and informed the distribution placed on this value in the probabilistic sensitivity analyses.

FIGURE 62 *Proportion of participants discontinuing treatment by time from start of treatment (weeks) from the RCT data*



Each solid line represents a single RCT. The dotted line in each plot is the estimate used in the base-case analyses.

FIGURE 63 *Assumed pattern of treatment discontinuation for all drugs (base-case in bold line and sensitivity analyses)*



7.3.8. Health state occupancy

IPD from the study by Wolstenholme and colleagues[181] was used to estimate the proportion of the total cohort in each of the following three health states at each cycle of the model (1 month): pre-institutionalized, institutionalized, and dead. A summary of the data from Wolstenholme and colleagues was given in Section *7.3.3*. For all 92 participants in the IPD, the median and mean time to end of pre-institutionalization was 1.8 and 2.4 years respectively (see *Figure 64*). The median and mean overall survival was 2.7 and 3.3 years respectively (see *Figure 65*). To calculate an equation representing time to end of pre-institutionalization, an exponential survival regression model ("survreg" routine from the "survival" "R" package) was fitted, with time to end of pre-institutionalization as the response variable, and MMSE, Barthel-ADL and age at the start of study as covariates. Note that the phrase 'time to end of pre-institutionalization' is not quite the same definition as time to institutionalization since some individuals died before entering institutionalization. For simplicity, the exponential distribution was chosen, rather than more complex two-parameter functions. Age was found to be a highly statistically significant predictor of time until end of

pre-institutionalization. Although MMSE and Barthel-ADL were not identified as statistically significant variables in explaining the variance of time to end of pre-institutionalization, both were retained in the model so that a treatment effect could be incorporated into the decision model. In the deterministic case, the time to pre-institutionalization was described by an exponential distribution with the following rate parameter:

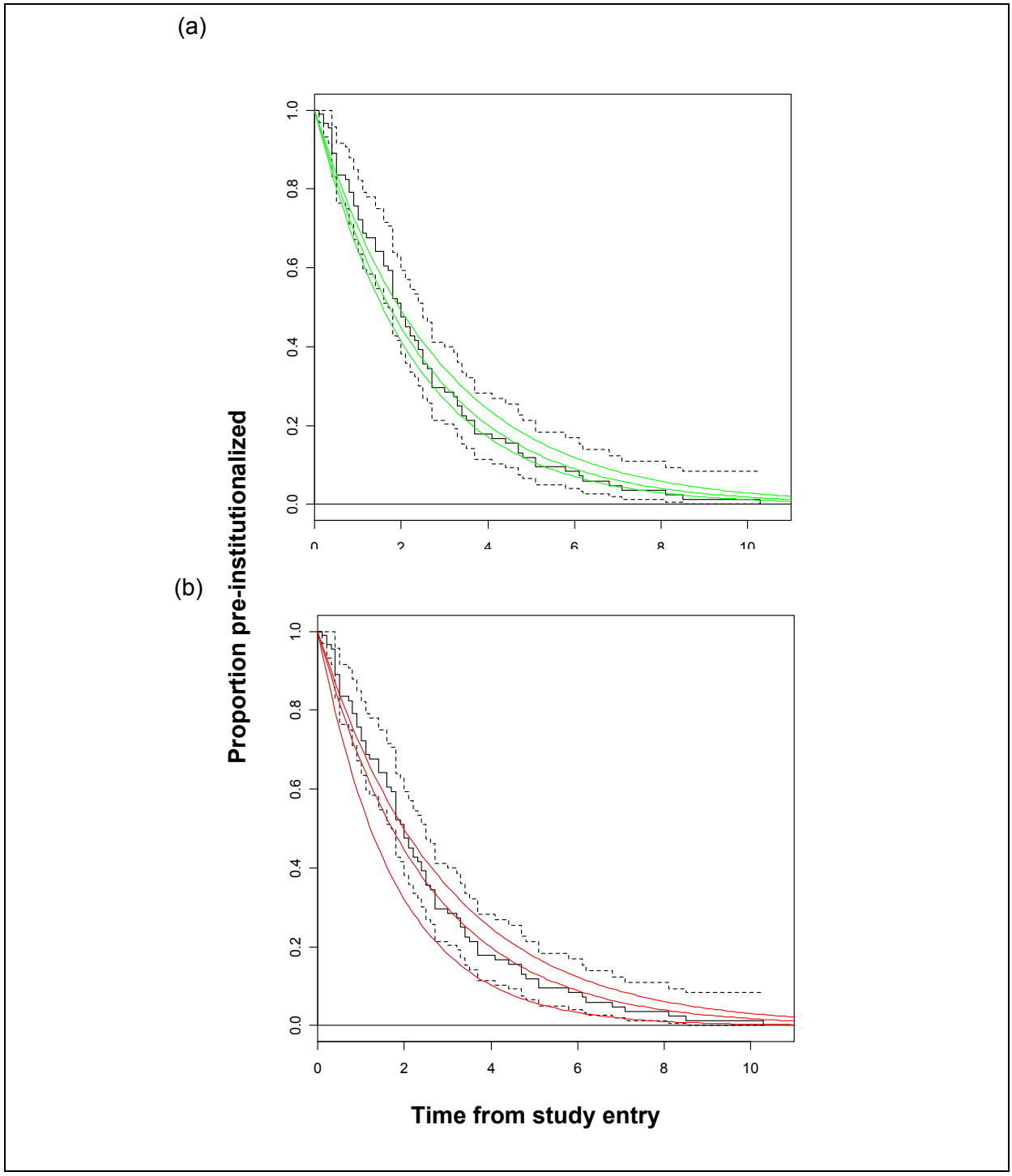
$$\lambda_{Pre-inst} = 1 \Big/ \exp\left(\begin{array}{l} 4.928 + 0.00409 * \text{mmse at study entry} + \\ 0.02139 * \text{ADL at study entry} - 0.05735 * \text{age at study entry} \end{array}\right)$$

As expected, the greater the MMSE, the greater the Barthel-ADL, and the lower the age of the individual at study entry, the longer that individual remained pre-institutionalized (*Figure 64*). For the PSA, it is necessary to model the uncertainty in the covariate coefficients. This was achieved as follows. First, the Cholesky matrix **C**, corresponding to the variance/covariance matrix of the parameter coefficients was calculated as

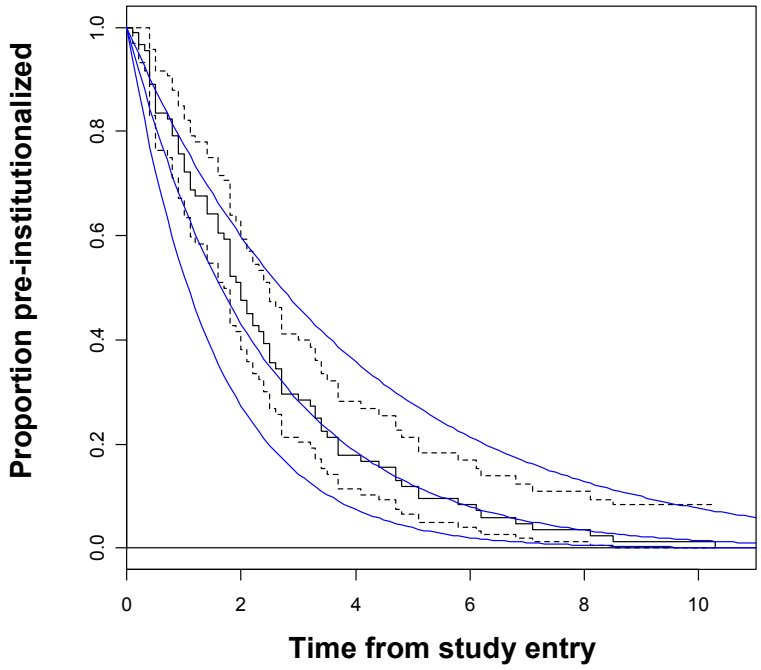
$$C = \begin{pmatrix} 1.6026 & 0 & 0 & 0 \\ -0.0015 & 0.0177 & 0 & 0 \\ -0.0159 & -0.0147 & 0.0237 & 0 \\ -0.0166 & -0.0000 & -0.0050 & 0.0014 \end{pmatrix},$$

where the rows and columns of **C** correspond to the intercept, MMSE at start of study, Barthel-ADL at start of study, and age at start of study respectively. Probabilistic covariate coefficients were then simulated as **y** + **Cz**, where **y** is the vector of coefficient means (given in the equation for $\lambda_{pre\text{-}inst}$ above), and **z** is a vector of independent standard normal variables[197].

FIGURE 64 *Proportion of cohort pre-institutionalized over time with model fit by (a) MMSE, (b) Barthel ADL and (c) age*



(c)



(a) lower curve MMSE = 10, middle curve MMSE = 14.5 (mean), upper curve MMSE = 20, (b) lower curve Barthel = 10, middle curve Barthel = 17 (mean), upper curve Barthel = 20, (c) lower curve age = 85, middle curve age = 78 (mean), upper curve age = 70

To model overall survival, an exponential distribution with the same covariates was fitted to the data with overall survival as the response variable. Overall survival was described by an exponential distribution with the following rate parameter;

$$\lambda_{OS} = 1 \Big/ \exp\left(\begin{array}{l} 4.322 + 0.00228 * \text{mmse at study entry} + \\ 0.04173 * \text{Barthel at study entry} - 0.04875 * \text{age at study entry} \end{array}\right)$$

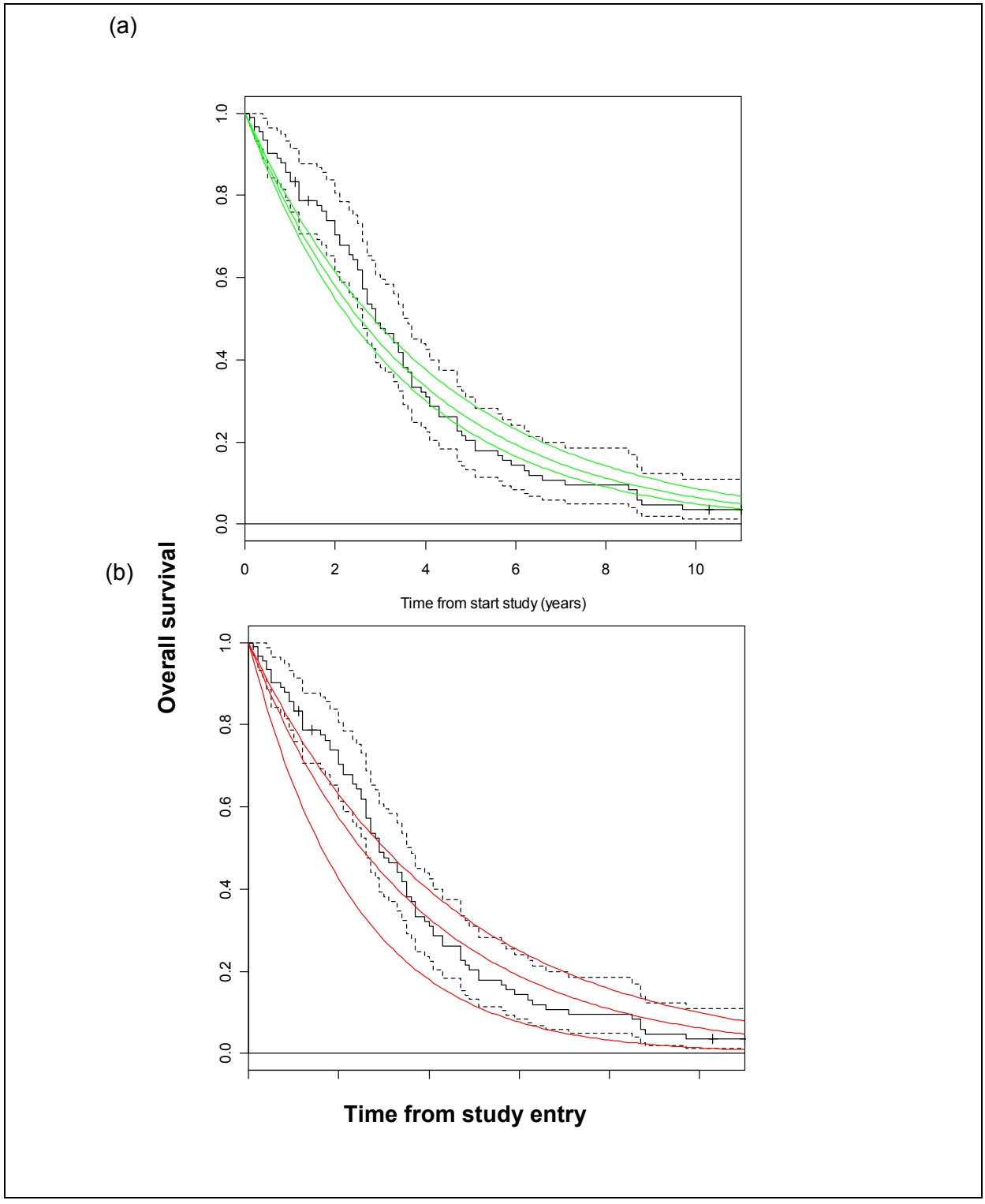
Again, as expected, the greater the MMSE, the greater the Barthel-ADL, and the lower the age of an individual at study entry, the longer that individual survived (see *Figure 65*). For the PSA, the Cholesky matrix was calculated as

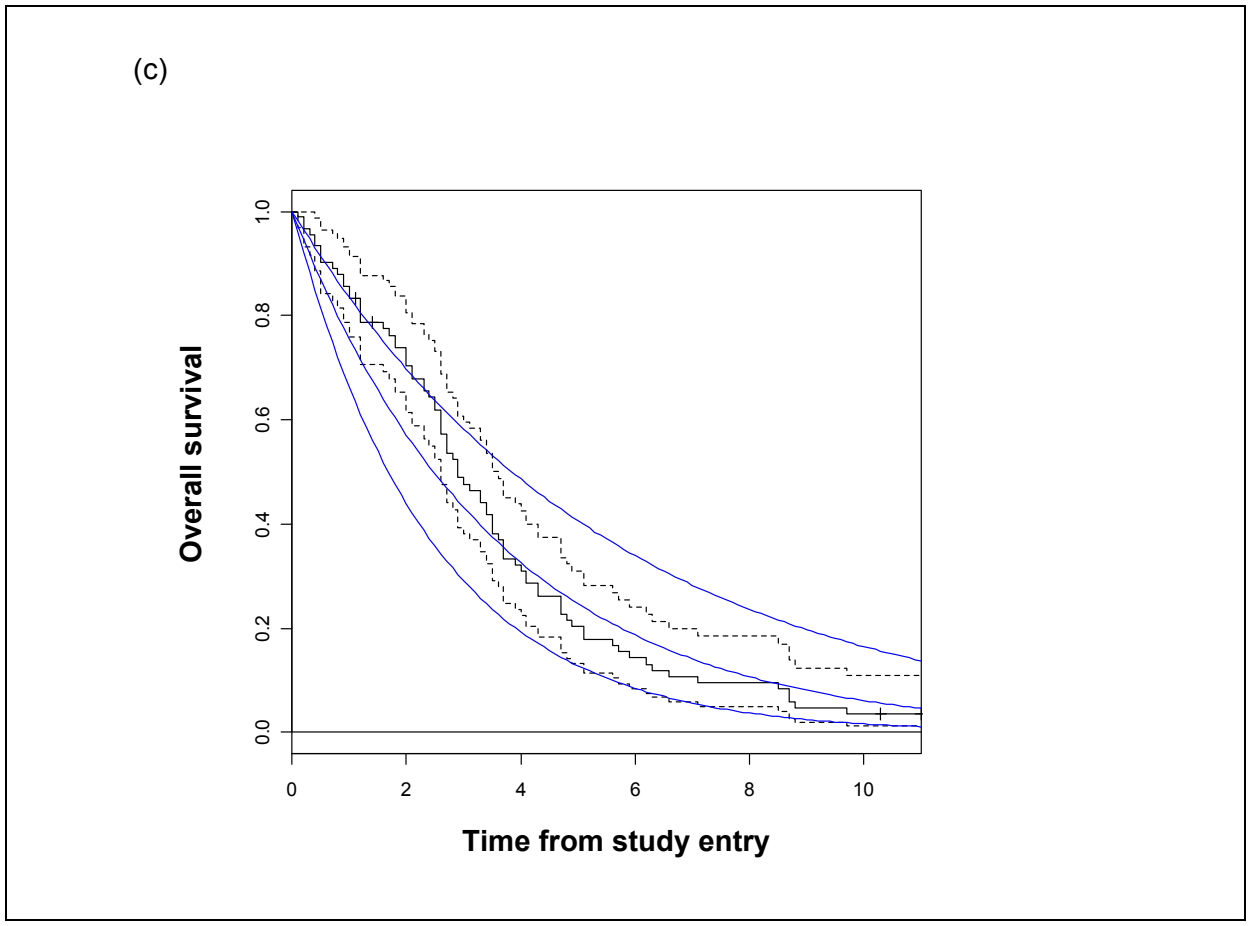
$$\begin{pmatrix} 1.5646 & 0 & 0 & 0 \\ -0.0005 & 0.0178 & 0 & 0 \\ -0.0160 & -0.0158 & 0.0253 & 0 \\ -0.0163 & 0.0002 & -0.0054 & 0.0014 \end{pmatrix}.$$

Probabilistic covariate coefficients were then simulated as before, using the vector of coefficient means given in the equation for λ_{OS} above. For the PSA, we assumed no correlation between the covariate coefficients for pre-institutionalisation and overall survival.

Thus, assuming exponential distributions with the above two rate parameters, the proportion of the cohort in any of the three states at any time period could be obtained. For the best supportive care cohort, the baseline age, MMSE and Barthel-ADL scores given in *Table 105* were inputted into these rate parameters, while for the treated cohorts after the six month initial treatment period the MMSE and ADL scores used were the baseline values plus the treatment effect.

FIGURE 65 *Proportion of cohort alive over time with model fit by (a) MMSE, (b) Barthel ADL and (c) age*





(a) lower curve MMSE = 10, middle curve MMSE = 14.5 (mean), upper curve MMSE = 20, (b) lower curve Barthel = 10, middle curve Barthel = 17 (mean), upper curve Barthel = 20, (c) lower curve age = 85, middle curve age = 78 (mean), upper curve age = 70

Transitions to institutionalization and death are both based on data from Wolstenholme and colleagues. It is likely that the individuals described in this study are not as ill as those in the general population for two important reasons. Firstly, inclusion criteria stated that individuals were not living in institutional care, and secondly individuals have had a diagnosis of Alzheimer's for a mean of 4.9 years.

7.3.9. Quality of life – utility estimates

In base-case analyses only utilities of people with Alzheimer's disease are considered. In sensitivity analyses, data concerning carer utility are also included. Evidence on the quality of life of people with Alzheimer's disease is reviewed and discussed in the next section, with the limited evidence available for carer's quality of life reviewed and discussed in Section *7.3.9.2*.

7.3.9.1. Quality of life of the individual with Alzheimer's disease

Since the previous review of the four drug treatments for Alzheimer's disease[2], a number of papers reporting utility values for people with Alzheimer's disease have been published. This literature is reviewed below (and summarised in *Table 110*), grouped by whether just one or multiple utilities are reported. The actual utility values from each of these studies are given in Appendix 18. For completeness and comparison, the utilities from studies reviewed in the previous HTA are also provided in Appendix 18.

One state health utility: having Alzheimer's disease

In an assessment of clinical and demographic correlates with utility scores, Miller and colleagues [199] report (carer-proxy) HUI:3 utilities for "having Alzheimer's" at baseline, 3, 6 and 9 months for up to 359 patients from the US. Patients had a mean MMSE of 15 (SD 5.8; range 4-29) at baseline. Miller and colleagues note that the utilities they report (0.184, range 0.29,1) are lower than those reported elsewhere using HUI-3 or other scales and suggest this may be due to a higher proportion of patients in their study having serious psychiatric symptoms compared to other studies. Miller and colleagues also point out that HUI-3 utility estimates are often found to be lower than those from alternative scales (i.e. HUI-2, QWB and EQ-5D) since there is a "greater emphasis on cognition" with HUI-3 than with the other scales. Approximately 60% of patients were on treatment with a cholinesterase inhibitor at baseline.

Naglie and colleagues [200] report self- and carer-proxy rated health state values for "having Alzheimer's" comparing the EQ-5D, QWB and HUI-3 in 60 people with Alzheimer's disease in Toronto. Participants had a mean MMSE of 18.9 (SD 4.5). Naglie and colleagues [200] point out that their findings (ranging from a utility of 0.86 from patients to 0.42 from carer-proxies) may not be generalisable to all patients with Alzheimer's disease as only those able to complete two facilitated interviews were included in this study.

Multiple-state health utility: by cognition

Jonsson and colleagues [201] provide EQ-5D health state utilities by MMSE ranges, based on self-ratings and carer proxy-ratings for 208 people with Alzheimer's disease in Sweden, Denmark, Finland and Norway. The utilities are reported where both patient and carer-ratings are available, or where only patient or only carer ratings are available. These utilities,

ranging from 0.21 to 1, are from a prospective observational study, with 71% of patients receiving cholinesterase inhibitors at baseline.

Community-based time trade-off (TTO) health state values were elicited for dementia (not specifically Alzheimer's disease) by Ekman and colleagues [202] from members of the Swedish public aged 45-84 years old. Ekman and colleagues report the TTO values by age group (45-54, 55-64, 65-74 and 75-84 years) and sex of the members of public contributing to the utility values. The TTO values are reported by severity based on the CDR. Ekman and colleagues [202] found that age, gender and self-assessed health status were associated with the utilities elicited, however there was no consistent pattern across severity, e.g. age was only a significant factor for utilities elicited for patients with mild cognitive impairment and mild dementia. The questionnaire from which the TTO values were obtained was validated by colleague and participant comments. Four vignettes based on the CDR scale were given. The response rate for the TTO section of the study was just 30%, with participants stating that "it was impossible to imagine living with dementia". This study was funded by Novartis.

Multiple-state health utility: by cognition and dependency

In an examination of the relationship between MMSE, IADL and QoL in 100 people with Alzheimer's disease from elderly care centres in Australia, Wlodarczyk and colleagues [203] report mean utility from the self-reported AQoL (Assessment of Quality of Life) by MMSE and IADL for self- and carer-proxy ratings. This sample is a subset of patients from a global donepezil trial where patients received open-label treatment daily for 24 weeks. The "weights from the AQoL are determined using TTO from a weighted sample of Victorian population, designed to ensure representativeness of the Australia population".

Multiple-state health utility: by cognition, dependency and residential status

Andersen and colleagues [204] mapped answers from health status and ADL questions from a cross-sectional survey of 244 demented patients (67.2% of which had Alzheimer's disease) in Denmark to EQ-5D. This study was reviewed in the previous MTA although was then unpublished. Issues with the mapping of these answers to EQ-5D have been highlighted elsewhere[145], including, as pointed out by the authors themselves, that "questions in the study included an aspect of time that the EQ-5D does not". Health state utilities are reported

by severity (MMSE), dependency and residential setting. The interviews were undertaken by a nurse in the participant's home where a family or professional carer either helped answer the questions or later verified them.

Karlawish and colleagues have assessed both self-ratings [205] and carer-proxy ratings [206] for health status values using EQ-5D and HUI-2 for people with Alzheimer's disease. The patient-rated utility values were based on 93 respondents with a mean MMSE of 21.3 (SD 4.3). Only patients with a MMSE score > 11 who were not in a nursing home and had an identifiable carer were included in the study which was based in the US. Utility values are reported by MMSE, modified MMSE, QoL-AD, IADL, BADL, GDS and SF-12 from the EQ-5D and HUI-2. Results for the carer-proxy ratings of health state utilities are based on responses from 100 carers, again reported by MMSE, modified MMSE, QOL-AD, IADL, BADL and SF-12. Only utilities derived from patient self-ratings and carer-proxy ratings by MMSE, IADL and BADL are reported in the table below.

TABLE 110 *Summary of the evidence providing utility values for individuals with Alzheimer's disease*

Source	Sample	Scale	Categories	Comments
Kerner et al [208]	Alzheimer's disease 159 Alzheimer's disease Spousal carer-proxy US	QWB	Alzheimer's disease patients Controls	
Miller et al [199]	Alzheimer's disease Up to 359 Carer-proxy US	HUI-3	Baseline 3 months 6 months 9 months	60% of patients on cholinesterase inhibitors
Naglie et al [200]	Alzheimer's disease 60 Self and carer-proxy Toronto	EQ-5D QWB HUI-3		Only patients able to complete two interviews are included
Jonsson et al [201]	Alzheimer's disease 208 Self and carer-proxy Sweden, Denmark, Finland, Norway	EQ-5D	MMSE: 26-30, 21-25, 15-20, 10-15, 0-9	
Sano et al	Alzheimer's disease Alzheimer's disease experts & students US	TTO & VAS	CDR: 1 & 3	
Ekman et al	Dementia General public (45-84 yrs) Sweden	TTO	CDR: 0.5, 1, 2, 3	Response rate of only 30%.
Neumann et al	Alzheimer's disease 679 carer-proxy US	HUI-2	CDR: 0.5, 1, 2, 3, 4, 5	
Wlodarczyk et al [203]	Alzheimer's disease 100 residents self- and carer-proxy Australia	AQoL	MMSE: 0-10, 10-15, 15-20, 20-25, 25+ IADL: 0-2, 3-5, 6-8	Part of global donepezil trial
Andersen et al [204]	Dementia 244 combined carer & patient answers Denmark	EQ-5D	MMSE: 21+, 10-20, 0-9 Dependency: Independent/ dependent Residential status: Community/ institution	Mapped from health status questionnaire and ADL answers
Karlawish et al [206;207]	Alzheimer's disease Self- and carer-proxy US	EQ-5D HUI-2	MMSE: 24-29, 20-23, 11-19 IADL: 8-10, 11-14, 15-27 BADL: 6, 7-14	Utilities also reported by other measures: 3MS, QoL-AD, GDS & SF-12

From the above review it is clear that a decision between the use of patient self-rated quality of life and carer-proxy rated quality of life is needed. Differences in utilities derived from patient's own or carer proxy based quality of life ratings have been found in a number of studies of people with Alzheimer's disease[200;201;203;205;206]. Such differences have been noted in other disease areas. For dementia and Alzheimer's disease authors have noted that these differences may be explained by carer

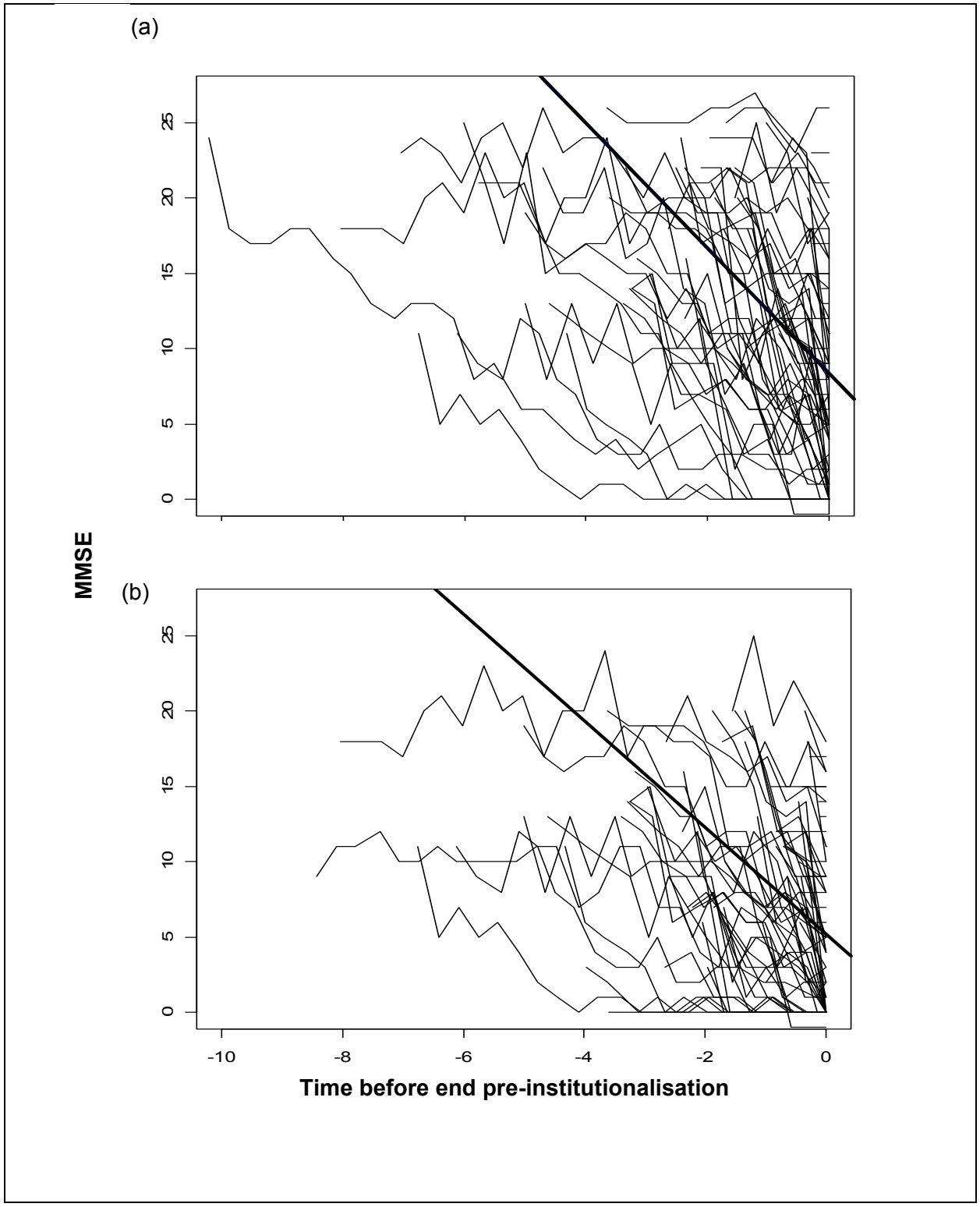
depression and/or burden (Karlawish and colleagues[208] and Sands and colleagues [209] as cited in Vogel and colleagues[210]) or lack of insight on the part of the patient [210]. There is no evidence that for people with moderate to severe Alzheimer's disease, reliable and consistent self-ratings of utility can be obtained. For example, Jonsson and colleagues[201] report self-rated utilities of 1 and 0.94 in patients with MMSE 10-15 and MMSE < 10, respectively. Although some may caution the use of carer proxy ratings [211], given the inconsistent quality of life observed with self-ratings of people with Alzheimer's disease, the base-case analysis uses carer-proxy ratings. In sensitivity analyses, results using patient self-reports are provided.

None of the studies reporting utilities have been carried out in the UK and only four have used the EQ-5D tool preferred in economic evaluations by NICE [195]: Andersen and colleagues[204], Karlawish and colleagues[205;206] and Jonsson and colleagues[201]. Note that Andersen and colleagues [204] mapped responses to general questionnaires to EQ-5D which involves some uncertainty as to the appropriateness of the mapping and the meaningfulness of the subsequent EQ-5D values.

With regard to utility, a particular criticism concerning the SHTAC model was that pre-full time care was too heterogeneous a state for a single utility (see number 11 in Appendix 17). Use of the IPD from Wolstenholme and colleagues provided a relationship between MMSE and time to end of pre-institutionalization. Since a number of sources have reported utility values by MMSE it is possible to map time prior to the end of pre-institutionalization to utility.

Within the IPD from Wolstenholme and colleagues, MMSE is recorded for each patient at multiple follow-ups from study entry, and so is a repeated measures dataset. A linear mixed effects model (from the "nlme" "R" package) was fitted with MMSE as the response variable, time to end of pre-institutionalization as a fixed effect, and patient as a random effect. Variation in the intercept and slope across patients was modelled as normal distributions. A fixed effect variable indicating whether a patient had mild to moderate Alzheimer's disease at the start of the study, or moderate to severe Alzheimer's disease was included in the model. Assuming t=years before the end of pre-institutionalization, as above, for patients with mild to moderate Alzheimer's disease, the following equation was obtained; MMSE = 8.34 + 4.17t. The corresponding equation for patients with moderate to severe Alzheimer's disease was MMSE = 5.18 + 3.55t (see *Figure 66*). Note that higher order terms for t were not modelled as they explained little of the variance in MMSE.

FIGURE 66 *MMSE as a function of time until end of pre-institutionalization with model fit for (1) mild to moderate AD and (b) moderate to severe AD*



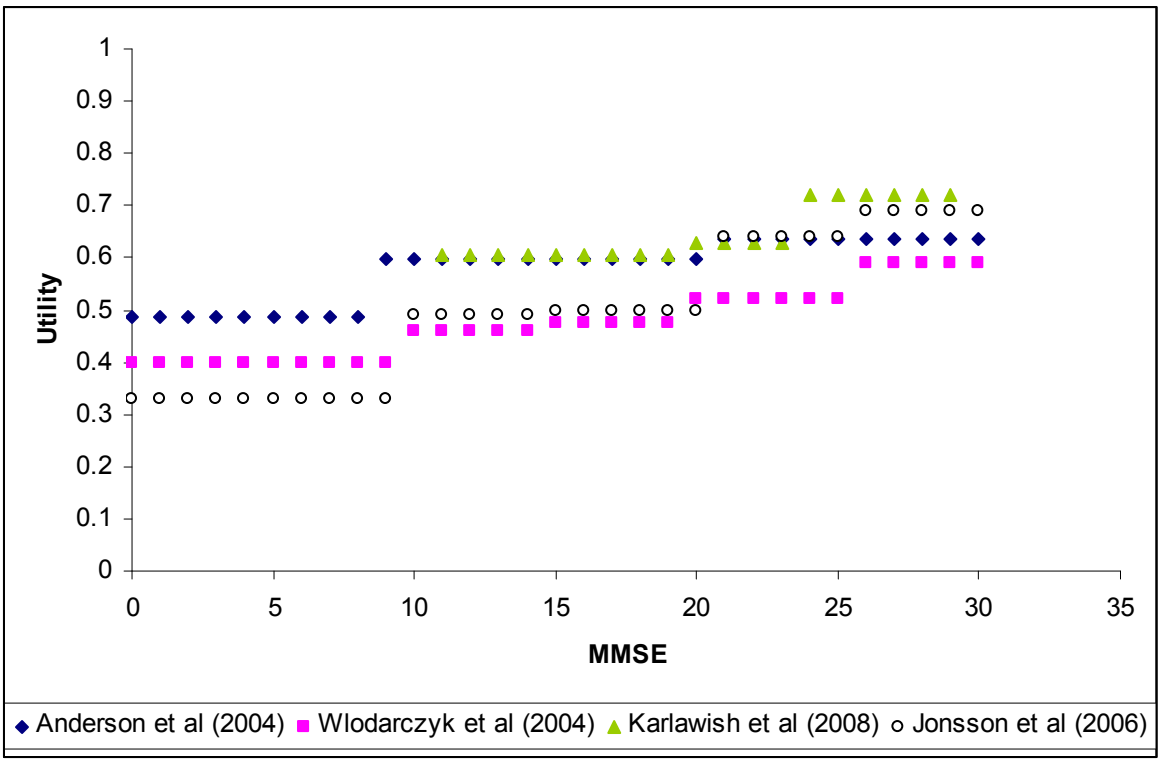
For the PSA, it was necessary to model the uncertainty in the covariate coefficients. The Cholesky matrix **C**, corresponding to the variance/covariance matrix of the parameter coefficients was calculated as

$$C_1 = \begin{pmatrix} 1.3805 & 0 & 0 & 0 \\ -0.1971 & 0.6227 & 0 & 0 \\ -1.3805 & 0.0000 & 0.7828 & 0 \\ 0.1971 & -0.6227 & -0.1062 & 0.3510 \end{pmatrix}, C_2 = \begin{pmatrix} 1.4621 & 0 & 0 & 0 \\ -0.1383 & 0.6129 & 0 & 0 \\ -1.4621 & 0.0000 & 0.8092 & 0 \\ 0.1383 & -0.6129 & -0.0881 & 0.3915 \end{pmatrix}$$

where C_1 corresponds to mild to moderate Alzheimer's disease and C_2 corresponds to moderate to severe Alzheimer's disease. The rows and columns of C_1 and C_2 correspond to the intercept, time to institutionalisation, indicator for mild to moderate or moderate to severe Alzheimer's disease to add to the intercept, and indicator for mild to moderate or moderate to severe Alzheimer's disease to add to the gradient respectively. Probabilistic covariate coefficients were then simulated as **y** + **Cz**, where **y** is the vector of coefficient means (given in the deterministic equations above), and **z** is a vector of independent standard normal variables[197].

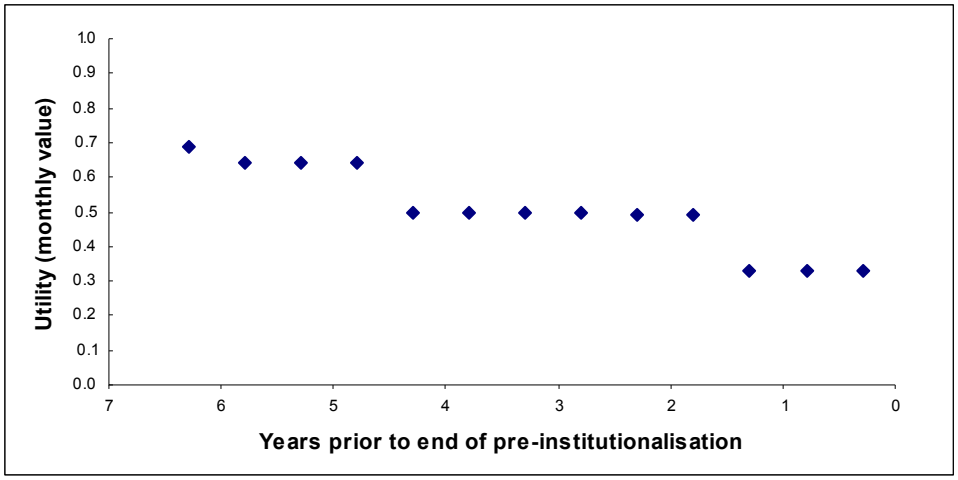
Utilities corresponding to these MMSE scores were required. Utility data by MMSE was available from five published studies[201;204;205;206;212], two of which considered the same population of people with Alzheimer's disease but reported patient self-ratings[205] and carer-proxy ratings[206] separately. The utility weights based on the carer-proxy ratings from these four studies are shown in *Figure 67*. The only study reporting direct EQ-5D valuations of utility across all MMSE scores was that by Jonsson and colleagues [201]. The utility weights from Jonsson and colleagues range from 0.69 for MMSE 30-26 to 0.33 for severe Alzheimer's disease. The evidence suggests that the utility weights from Jonsson and colleagues are not particularly different to those from the rest of the literature. Interestingly, the utility weight for MMSE<10 from Jonsson and colleagues is very similar to the utility weight for people with Alzheimer's disease who are defined as dependent from the study by Andersen and colleagues[204]: 0.33 from Jonsson and colleagues and 0.343 from Andersen and colleagues.

FIGURE 67 Carer proxy utility weights for people with Alzheimer's disease by MMSE



Mapping utility weights from Jonsson and colleagues[201] onto time until the end of pre-institutionalization from the equations above, shows that, as expected, the shorter the time until the end of pre-institutionalization, the lower the utility weight (see *Figure 68*). The utility values used in the base-case analysis are shown in *Table 111*. Note that utility in the institutionalized state is the same as for MMSE<10 since, as noted above, individuals have, on average, an MMSE<10 before being institutionalized.

FIGURE 68 *Plot of utility from Jonsson and colleagues[201] by time to end of pre-institutionalization used in the base-case analysis*



In sensitivity analyses the patient's self-rated utility from Jonsson and colleagues [201] were assessed for their impact on the cost–utility findings. Similarly, utility estimates from the AQoL utilities as reported by Wlodarczyk and colleagues[203] were assessed in one-way sensitivity analyses.

TABLE 111 *Utilities used in the base-case analysis (from Jonsson and colleagues[201]) and sensitivity analyses*

Health state	Value	N	PenTAG estimates of SD	Patient rated quality of life
Pre-institutionalization by MMSE				
0-9	0.33	44	0.151	0.78
10-14	0.49	88	0.107	0.73
15-20	0.5	83	0.110	0.83
21-25	0.49	25	0.200	0.85
26-30	0.69	22	0.213	0.84
Institutionalization (MMSE 0-9)	0.33	44	0.151	0.78
Dead	0			

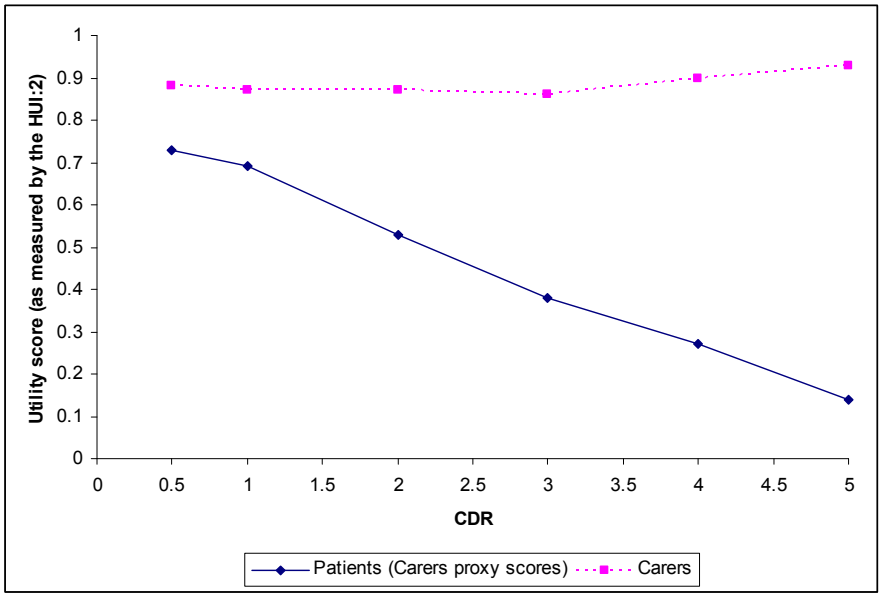
Note that no estimates of uncertainty are reported by Jonsson and colleagues[201] only the number of carers contributing to the mean estimate

For the PSA, estimates of the uncertainty of the utility values from Jonsson and colleagues were obtained by assuming that the standard deviation of these values was equivalent to $1/\sqrt{N}$. Although this is not ideal, it produced figures of the same magnitude as the standard deviations reported by Andersen and colleagues[204] for their utility estimates.

7.3.9.2. Quality of life of the carer

Very little data on the utility of the carer's of people with Alzheimer's disease was identified from the literature. A study by Neumann and colleagues[213], which was reviewed in the previous MTA, contained some evidence regarding carer utility (measured on the HUI:2 scale) by patient progression (measured on the CDR scale). This evidence suggests that the carer's utility scores remain fairly stable until the patient's disease progresses to a score of 3 on the CDR (indicating severe Alzheimer's disease) when the carer's quality of life starts to improve (see *Figure 69*). This is most likely due to the patient being placed in institutionalized care. However, note that this was a cross-sectional study that did not follow patients and their carer's over time, so may not be fully representative of carer's utility as the patient progresses.

FIGURE 69 *HUI:2 utility scores of patients (carers providing proxy scores) and carers by Clinical Dementia Rating (CDR) from a cross-sectional study in the US*



To include data on carer's utility in the PenTAG cost–utility model, the CDR scale was mapped onto the MMSE scale using the findings of a study by Perneczky and colleagues[214]. Perneczky and colleagues demonstrated that the CDR scale can be used to map onto MMSE ranges for people with dementia. Their analysis indicated substantial agreement (as measured by Cohen's kappa, κ) between the MMSE ranges and CDR stages for the mild, moderate and severe stages of dementia ($\kappa > 0.6$). These results are reproduced below in *Table 112.*

TABLE 112 *Mapping of CDR stages to MMSE scores from Perneczky and colleagues[215] and associated carer's utilities from Neumann and colleagues[214]*

CDR stage	MMSE range	Cohen's kappa, κ	HUI:2 utility
0 – No dementia	30	0.44	-
0.5 – Questionable dementia	26-29	0.28	0.88
1 – Mild dementia	21-25	0.62	0.87
2 – Moderate dementia	11-20	0.69	0.87
3 – Severe dementia	0-10	0.76	0.86

Assuming these mappings in *Table 112*, the carer's utility by the patient's MMSE score was obtained and then, as in Section *7.3.9.1*, mapped onto time prior to institutionalization for the person with Alzheimer's disease.

7.3.10. Cost estimates

The costs considered in the decision model are those that fall on the NHS and PSS. They are the drug costs, monthly costs of care (pre-institutionalized and institutionalized) and the costs of a 6-monthly monitoring outpatient visit for those treated with donepezil, galantamine, rivastigmine or memantine. It is assumed that any adverse events are mild, do not require further treatment, and so do not induce further costs. All costs are for 2009 to avoid incorporating further uncertainty for the inflation to 2010 costs where such costs are not yet published (e.g. Unit Costs of Health and Social Care (2009) and NHS Reference Costs (2008-9)). Note that the relevant drug costs do not differ between BNF58 (4th quarter 2009) and BNF59 (1st quarter 2010). A review of evidence on the costs associated with Alzheimer's disease did not identify data on the NHS and PSS costs of carer's of people with Alzheimer's disease. Therefore, no such carer's costs were included in the PenTAG decision model.

7.3.10.1. Drug costs for Alzheimer's disease

Monthly drug costs were calculated from costs reported in the BNF 58 for the specific doses of interest. For galantamine, rivastigmine capsules and memantine, a mix of doses were assumed which is reflected in the monthly costs presented in *Table 113*. It was also assumed that treated individuals would have a six-monthly out-patient monitoring visit. The cost of such a visit was obtained from the National Schedule of Reference Costs 2008-9.

All costs used in the base-case analysis are shown in *Table 113*. As noted in Section *7.3.7*, the drug costs and associated 6-monthly out-patient costs are adjusted in the model to reflect treatment discontinuation by a proportion of the treated cohort.

***TABLE 113** Drug and NHS and PSS care costs used in the base-case analysis*

Cost component		£ (2009)	Source
Drugs (monthly cost)	Donepezil (10mg/day)	£97	Calculated from the daily drug cost of £3.18 (BNF58)
	Galantamine (16-24mg/day)	£83	Calculated as an average of daily costs for 16mg (£2.44) and 24mg (£3.66), leading to a daily cost of £2.72 (BNF58)
	Rivastigmine capsules (9-12mg/day)	£98	Calculated as a weighted average of daily costs for 9mg (0.7*£3.56) and 12mg (0.3*£2.37), leading to daily drug cost of £3.21
	Rivastigmine patches ($10cm^2$)	£79	Calculated from the daily cost of £2.60
	Memantine (15-20mg/day)	£71	Calculated as a weighted average of daily costs for 15mg (0.2*£1.85) and 20mg (0.8*£2.46), leading to daily drug cost of £2.34
Outpatient visit	6-monthly visit	£158	National Schedule of Reference Costs (2008-9). NHS Trusts consultant led; Follow-up attendance non-admitted face-to-face
	Monthly cost	£26	
NHS & PSS	Pre-institutionalized	See equation *(12)* and equation *(13)*, below	IPD from Wolstenholme et al [181]
	Institutionalized	£2,941	

7.3.10.2. Cost of health and social care received by Alzheimer's disease patients

A review of published research evidence failed to identify up-to-date estimates of the NHS and PSS costs associated with Alzheimer's disease. For the PenTAG model it was therefore necessary to use available data from a number of years ago and inflate the relevant costs as appropriate. Data from the Oxfordshire cohort study by Wolstenholme and colleagues[181] was used to provide NHS and PSS costs for the PenTAG model. Information on the following

resource use was recorded by Wolstenholme and colleagues at baseline and each subsequent follow-up interview:

- Number and duration of acute hospitalisations and respite care
- Number of outpatient visits
- Day care and home attendances by district nurses, community psychiatric nurses, home helps or other care assistants
- Number of visits by or to the GP or practice nurse.

All of these items of service use were recorded whether they were related to Alzheimer's disease or other health problems. Data on the use of wheelchairs, bath or bed hoists and incontinence pads and sheets were recorded. The individual's current place of residence was also noted so that accommodation costs could be calculated. Wolstenholme and colleagues report that unit costs were taken mainly from the Unit Costs for Health & Social Care for 1998 and 1999, supplemented by hospital trust financial returns and the Chartered Institute of Public Finance and Accounting in addition to data from surveys of local hospitals, and from specific residential and nursing homes, carried out by Wolstenholme and colleagues[181] (see Table 1 of Wolstenholme and colleagues[181]).

To be used in the PenTAG model the monthly costs calculated by Wolstenholme and colleagues (implicitly 1998 costs) were inflated to 2009 costs using an inflation factor of 1.54 (= 267.0/173.5) from the inflation indices for hospital and community health services in Unit Costs of Health & Social Care annual reports for 2009 and 2004)[215;216]. Pre-institutionalisation costs were all assumed to fall on the NHS or PSS budget, but post-institutionalisation a proportion of accommodation costs were assumed to be self-funded (i.e. paid by patients or their families). On the basis of data in the Dementia UK report (as cited by the 2007 NAO report[5;6]), we assume that being in residential care[6] is equivalent to being institutionalized in our model, and 94% of the non-informal care costs of being institutionalised are accommodation (or "care home costs"). Of these, 30% were reported in 2007 to be self-funded (i.e. not NHS or social services department). So, in our base-case analysis we assume that 28% of post-institutionalisation costs (excluding informal care costs) are self-funded, and 72% are NHS and PSS funded.

A criticism of the SHTAC model that was explored in the PenTAG model was that the pre-full time care health state was too heterogeneous to represent a single NHS and PSS cost (see number 5 in Appendix 17). The IPD from Wolstenholme and colleagues allowed an exploration of this criticism by developing a relationship between the inflated cost per month and the time before the end of pre-institutionalization. See Appendix 19.

A linear mixed effects model with the inflated cost per month as the response variable, a cubic equation for the time to end of pre-institutionalization as a fixed effect (higher order terms were non-significant), and patient as a random effect, where variation in the intercept across patients was modelled as a normal distribution. An indicator explanatory variable was included for whether a patient had mild to moderate Alzheimer's disease at the start of the study, or moderate to severe Alzheimer's disease.

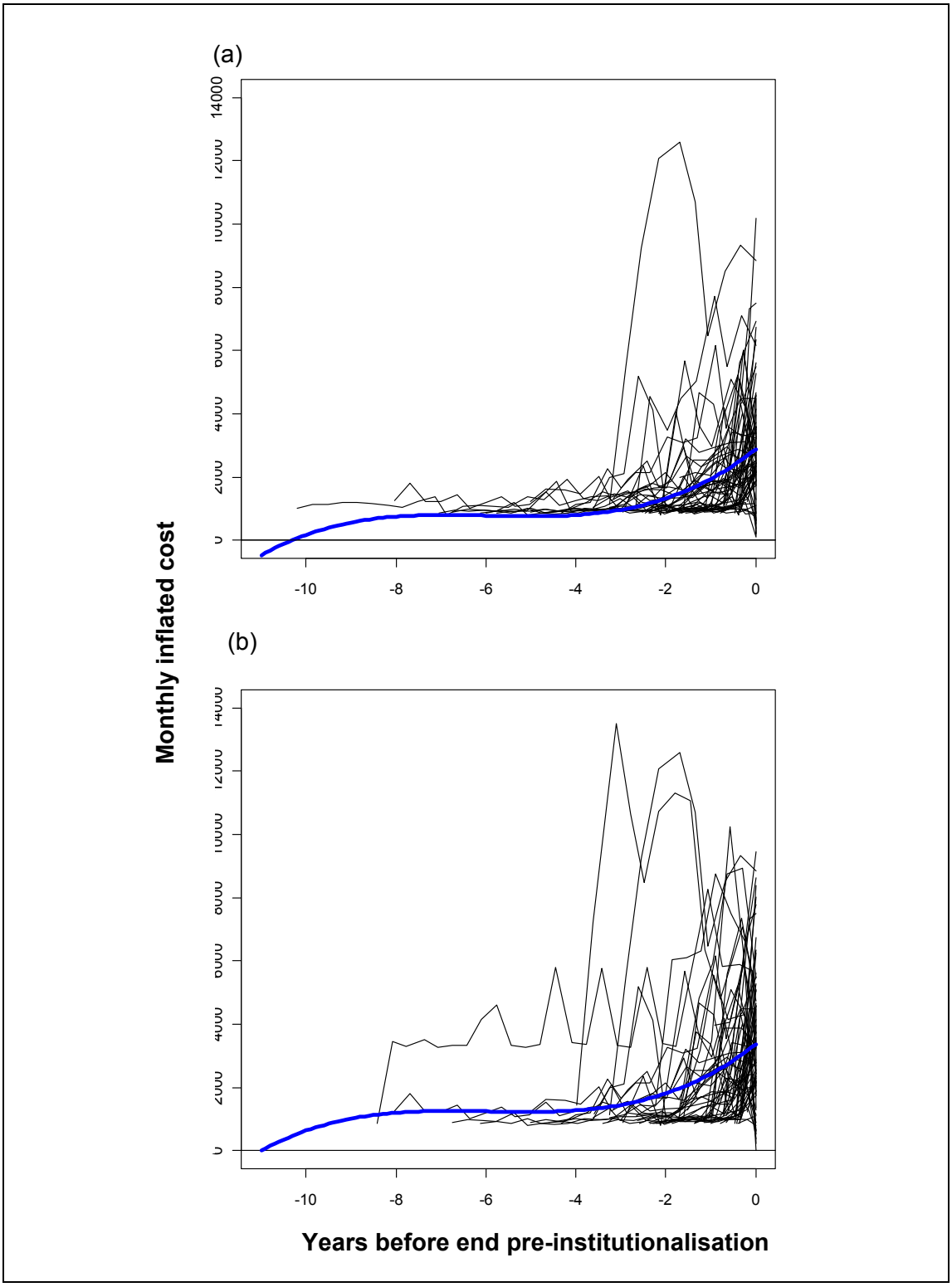
For patients with mild to moderate Alzheimer's disease, the following equation was obtained (see *Figure 70*a):

$$\text{Monthly inflated cost (£)} = 2877 - 1122t + 194t^2 - 10.9t^3, \quad \textbf{(12)}$$

where t = years before the end of pre-institutionalization. The corresponding equation for patients with moderate to severe Alzheimer's disease was (see *Figure 70*b)

$$\text{Monthly inflated cost (£)} = 3363 - 1117t + 191t^2 – 10.7t^3. \quad \textbf{(13)}$$

FIGURE 70 *Monthly inflated cost as a function of time until pre-institutionalization showing model fit for (a) mild to moderate AD and (b) moderate to severe AD*



For the PSA, it is necessary to model the uncertainty in the covariate coefficients, and this was done by calculating the Cholesky matrix **C**, corresponding to the variance/covariance matrix of the parameter coefficients, as

$$C_1 = \begin{pmatrix} 325.6 & 0 & 0 & 0 & 0 \\ -36.97 & 147.69 & 0 & 0 & 0 \\ 9.53 & -43.65 & 17.50 & 0 & 0 \\ -0.66 & 3.23 & -2.05 & 0.59 & 0 \\ -303.71 & -76.08 & -32.99 & -10.79 & 178.33 \end{pmatrix}, C_2 = \begin{pmatrix} 343.01 & 0 & 0 & 0 & 0 \\ -40.30 & 147.28 & 0 & 0 & 0 \\ 9.79 & -43.71 & 17.47 & 0 & 0 \\ -0.66 & 3.23 & -2.05 & 0.59 & 0 \\ -319.50 & -71.83 & -27.87 & -7.13 & 186.91 \end{pmatrix}$$

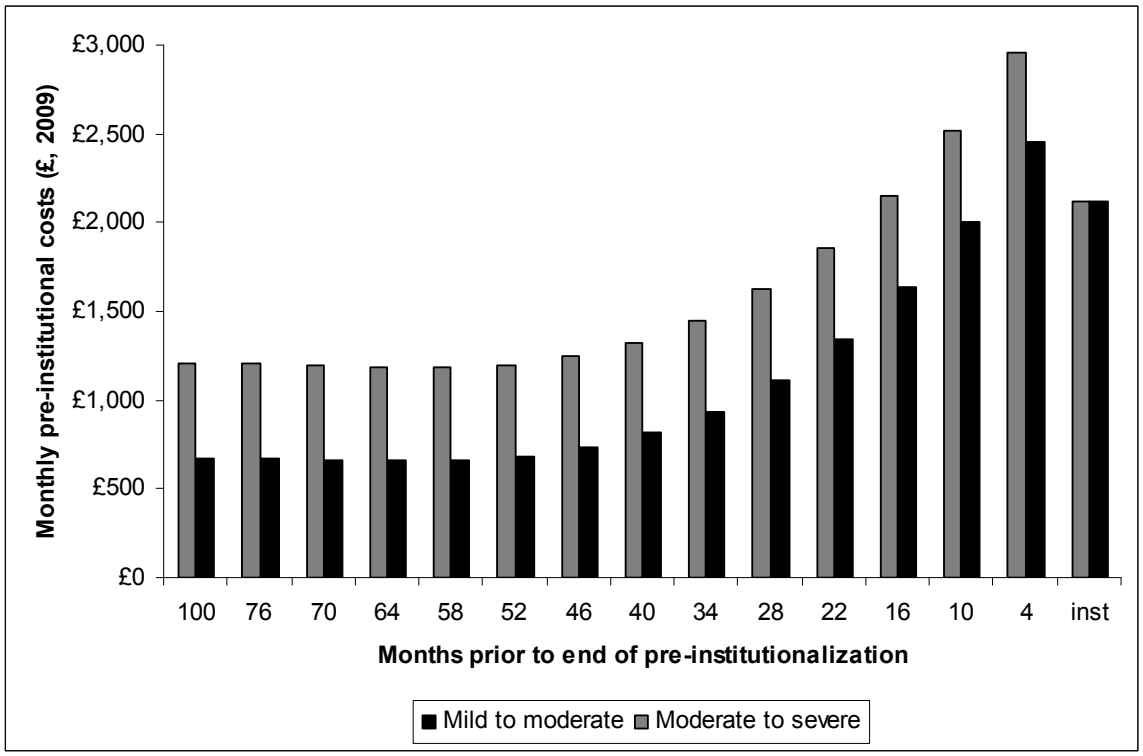
where C_1 corresponds to mild to moderate Alzheimer's disease and C_2 corresponds to moderate to severe Alzheimer's disease. The rows and columns of C_1 and C_2 correspond to the intercept, time to institutionalisation, square of time to institutionalization, cube of time to institutionalization, and indicator for mild to moderate or moderate to severe Alzheimer's disease to add to the intercept. Probabilistic covariate coefficients were then simulated as **y** + **Cz**, where **y** is the vector of coefficient means (given in the deterministic equations above), and **z** is a vector of independent standard normal variables [197].

Thus, at one year until the end of pre-institutionalization, the mean NHS and PSS cost per participant for mild to moderate participants was £1938 per month, while for moderate to severe participants the cost was £2427 per month. In the cost–utility model, at each cycle, the proportion of the cohort within 6-monthly time-periods of leaving the pre-institutionalized state was calculated. The time-periods were 0-6 months, 7-12 months, 13-18 months, and so on, until 72 months. The mid-points of these 6 monthly time-periods were used to calculate MMSE and costs prior to institutionalization. Any individuals >72 months prior to institutionalization were assumed to be 75 months prior to institutionalization. A plot of the monthly pre-institutionalized costs by time to institutionalization used in the PenTAG model is given in *Figure 71*.

To calculate a mean monthly cost while in institutional care, a linear mixed effects model was fitted, with the inflated cost per month as the response variable, fitting just the intercept term, with patient ID as a random effect. The overall cost was assumed independent of disease severity, and independent of time since institutionalized. Visual inspection suggested that the second point was approximately true. The overall mean monthly cost was £2,941. For the PSA, the monthly cost was assumed normally distributed with standard error of £180.

Only 72% of this monthly cost for institutional care is included in the PenTAG model to account for 28% of institutionalized costs being privately funded as described above. Thus, the monthly cost of institutional care is £2,117.

FIGURE 71 *Monthly costs (£, 2009) by time to institutionalization used in the base-case analyses*



Note that when costs for institutional care are compared to the pre-institutionalized costs (Final data point in *Figure 71*), it can be seen that institutional care costs are lower than the pre-institutionalized costs for mild to moderate individuals within 4 months of being institutionalized, and moderate to severe individuals within 16 months of being institutionalized. The total institutional care costs in the PenTAG model have been subjected to a 72% reduction to account for the fact that not all of the institutional care costs are NHS/PSS funded. The pre-institutionalized costs have not been subjected to this reduction. The pattern of care costs shown in *Figure 71* reflects increases in care costs to the NHS/PSS as individual progress to requiring institutionalization. Once in institutional care, fewer NHS/PSS costs are incurred.

7.3.10.3. Cost estimates used in sensitivity analyses

The drug costs reported by Lundbeck and Eisai in their industry submission are used in one-way sensitivity analyses, as are the industry estimates of the monthly cost of institutionalization and the six-monthly out-patient follow-up visit cost. The impact of assuming that 72% of the monthly £2941 cost of institutionalization are NHS/PSS funded is also assessed. It is assumed that 53% or 91% of institutionalised costs are NHS/PSS costs. Note that these percentages do not refer to the industry cited costs as these are assumed to have already accounted for the fact that not all institutionalization costs are funded by the NHS/PSS.

7.3.11. Key assumptions of PenTAG model

To summarise, key assumptions in the base-case analysis are that there was no treatment effect on survival, ADCS-ADL was transformed onto the Barthel ADL, time to institutionalization and overall survival in clinical practice are similar to that experienced in the Oxfordshire study by Wolsteholme and colleagues [181] and carer-proxy utility values were used for patient quality of life. The pre-institutionalized state allowed for a relationship between (i) utility and time prior to end of pre-institutionalization and (ii) cost and time prior to pre-institutionalization. Costs and utilities were assumed to be constant within the institutionalized state. All parameter values used in the base-case analyses for mild and moderate Alzheimer's disease and moderate to severe Alzheimer's disease are presented in *Table 114* and *Table 115*, respectively. Probabilistic sensitivity analysis was undertaken for both of the base-case analyses: mild to moderate and moderate to severe Alzheimer's disease. In both PSAs, distributions were placed on uncertain parameters. The distributions used are also presented in *Table 114* and *Table 115*, and shown graphically in Appendix 20. For each PSA, 10,000 simulations of the cost–utility model were undertaken.

TABLE 114 Parameter values used in the base-case analysis for individuals with mild to moderate Alzheimer's disease

Parameter	Value	Std err	Source	Justification	Distribution for PSA*
Cohort characteristics					
Mean age					
Group 1: 25% cohort	69	NA	IPD from Wolstenholme et al[181]	Based on data from a UK epidemiological study on which AD progression is modelled for mild to moderate severity of AD	NA
Group 2: 50% cohort	77	NA	IPD from Wolstenholme et al[181]		NA
Group 3: 25% cohort	86	NA	IPD from Wolstenholme et al[181]		NA
Mean MMSE	17	NA	IPD from Wolstenholme et al[181]	Based on data from the UK epidemiological study on which AD progression is modelled for mild to moderate severity of AD	NA
Mean Barthel ADL	17.52	NA	IPD from Wolstenholme et al[181]	Based on data from the UK epidemiological study on which AD progression is modelled for mild to moderate severity of AD	NA
Proportion starting model in institutionalized state	0.1		LASER-AD	Based on data from a UK epidemiological study, and evidence of a reduction in the number of individuals institutionalised for mild to moderate severity of AD	Beta(1,9)
Time horizon	20 years			Estimated that <5% of cohort are still alive	NA
Discounting costs	3.5%		NICE methods guide[195]	As stated in NICE methods guide	NA
Discounting benefits	3.5%		NICE methods guide[195]	As stated in NICE methods guide	NA
Clinical effectiveness					
Donepezil (10mg)					
MMSE	1.24	0.216	Meta-analysis result from Chapter 5	Based on data from a systematic review of the evidence	Normal(1.24, 0.216) From 95% CI of pooled estimate
ADCS-ADL	2.02	0.470	Average of value from galantamine and rivastigmine capsules	Since no evidence were identified, this is based on an assumption of a class effect	Normal(2.17, 0.470) From 95% CI of pooled estimate
Galantamine (16-24mg)					
MMSE	1.13	0.156	Average of value from donepezil and rivastigmine capsules	Since no evidence were identified, this is based on an assumption of a class effect	Normal(1.13, 0.156) From 95% CI of pooled estimate
ADCS-ADL	2.23	0.462	Meta-analysis result from Chapter 5	Based on data from a systematic review of the evidence	Normal(2.23, 0.462) From 95% CI of pooled estimate

Parameter	Value	Std err	Source	Justification	Distribution for PSA*
Rivastigmine capsules (9-12mg)					
MMSE	1.02	0.225	Meta-analysis result from Chapter 5	Based on data from a systematic review of the evidence	Normal(1.02, 0.225) From 95% CI of pooled estimate
ADCS-ADL	1.80	0.818	Single study	Based on only RCT reporting outcome	Normal(1.8, 0.818) From 95% CI of pooled estimate
Rivastigmine patches (10cm^2)					
MMSE	1.10	0.296	Single study	Based on only RCT reporting outcome	Normal(1.1, 0.296) From 95% CI of estimate
ADCS-ADL	2.20	0.808	Single study	Based on only RCT reporting outcome	Normal(2.2, 0.808) From 95% CI of estimate
Percentage of total cohort discontinuing treatment each month	4%		Based on mixture of evidence from RCTs and clinical opinion	Assumes most participants have discontinued treatment by 2 years (similar to AD2000[96] results)	Beta(12,290)
Health state utilities					
Pre-institutionalized (by MMSE)	Linear eqn with time		MMSE by time prior to pre-institutionalized calculated from IPD of Wolstenholme et al [181]		Cholesky matrix shown in Section *7.3.9.1*
MMSE: 0-9	0.33	0.044	Jonsson et al (2006)[201]	Utilities reported in Jonsson et al are similar to those reported by MMSE in other studies	Beta(36.59, 74.28)
MMSE: 10-14	0.49	0.039			Beta(78.04, 81.22)
MMSE: 15-20	0.5	0.012			Beta(856.27, 856.27)
MMSE: 21-25	0.64	0.011			Beta(1137.19, 639.67)
MMSE: 26-30	0.69	0.023			Beta(282.51, 126.92)
Institutionalized (MMSE 0-9)	0.33	0.044	Jonsson et al (2006)[201]	Analysis of IPD from Wolstenholme et al suggests participants had an average MMSE of 9 when 0.04 years prior to institutionalization, therefore institutionalization used as proxy for severe Alzheimer's disease	Beta(36.59, 74.28)
Dead	0				NA
Monthly drug costs					
Donepezil (10mg)	£97		BNF58		NA
Galantamine (16-24mg)	£83		BNF58		NA
% 16mg costs	50%			Author judgement on % RCT participants having mean dose of 16mg	%16mg=Normal(0.5,0.1)

Parameter	Value	Std err	Source	Justification	Distribution for PSA*
% 24mg costs	50%			or 24mg	% 24mg = 1 - % 16mg
Rivastigmine (≤12mg)	£98		BNF58		NA
% 9mg costs	70%			Author judgement on % RCT participants having mean dose of 9mg or 12mg	%9mg=Normal(0.3, 0.1)
%12mg costs	30%				%12mg = 1 - % 9mg
Rivastigmine patch	£79		BNF58		NA
6-monthly monitoring out-patient visit cost	£158		National Schedule of Reference Costs (2008-9).	NHS Trusts consultant led; follow-up attendance non-admitted face-to-face	Gamma(4.94, 32)
Monthly monitoring out-patient visit cost	£26		Calculated from 6-monthly cost of £158 for single visit		
Pre-institutionalized NHS/PSS costs	Cubic eqn with time		Relationship between NHS & PSS costs and time prior to institutionalisation calculated from IPD of Wolstenholme et al[181]		Cholesky matrix shown in Section *7.3.10.2*
Institutionalized	£2,941	£108	Calculated from IPD of Wolstenholme et al[181]	Monthly cost of institutional care. Only 72% of these costs are assumed to be NHS/PSS	Normal(2941, 108)
% institutionalized costs funded by NHS/PSS	0.72		Dementia UK report (as cited by the 2007 NAO report[5;6])	Assumes that 28% of institutional care costs are not funded by the NHS/PSS	Beta(15, 5.83)

* See Appendix 20 for graphical presentation of distributions

TABLE 115 *Parameter values used in the base-case analysis for individuals with moderate to severe Alzheimer's disease*

Parameter	Value	Std err	Source	Justification	Distribution for PSA*
Cohort characteristics					
Mean age					
Group 1: 25% cohort	69		IPD from Wolstenholme et al[181]	Based on data from a UK epidemiological study on which AD progression is modelled for moderate to severe AD	NA
Group 2: 50% cohort	78		IPD from Wolstenholme et al[181]		NA
Group 3: 25% cohort	87		IPD from Wolstenholme et al[181]		NA
Mean MMSE	11.73		IPD from Wolstenholme et al[181]	Based on data from a UK epidemiological study on which AD progression is modelled for moderate to severe AD	NA

Parameter	Value	Std err	Source	Justification	Distribution for PSA*
Mean Barthel ADL	16.34		IPD from Wolstenholme et al[181]	Based on data from a UK epidemiological study on which AD progression is modelled for moderate to severe AD	NA
Proportion starting model in institutionalized state	0.4		LASER-AD	Based on data from a UK epidemiological study, and evidence of a reduction in the number of individuals institutionalised for moderate to severe AD	Beta(4,6)
Time horizon	20 years			Estimated that <5% of cohort are still alive	NA
Discounting costs	3.5%		NICE methods guide[195]	As stated in NICE methods	NA
Discounting benefits	3.5%		NICE methods guide[195]	As stated in NICE methods	NA
Clinical effectiveness					
Memantine (20mg)					
MMSE	0.70	0.346	Meta-analysis result from Chapter 5	Based on data from systematic review of evidence	Normal(0.70, 0.346)
ADCS-ADL	1.41	0.70	Meta-analysis result from Chapter 5	Based on data from systematic review of evidence	Normal(1.41, 0.70)
Percentage of total cohort discontinuing treatment each month	4%		Based on mixture of evidence from RCTs and clinical opinion	Assumes most participants have discontinued treatment by 2 years (similar to AD2000[96] results)	Beta(12,290)
Health state utilities					
Pre-institutionalized (by MMSE)	Linear eqn with time		MMSE by time prior to pre-institutionalized calculated from IPD of Wolstenholme et al [181]		Cholesky matrix shown in Section *7.3.9.1*
MMSE: 0-9	0.33	0.044	Jonsson et al (2006)[201]	Utilities reported in Jonsson et al are similar to those reported by MMSE in other studies	Beta(36.59, 74.28)
MMSE: 10-14	0.49	0.039			Beta(78.04, 81.22)
MMSE: 15-20	0.5	0.012			Beta(856.27, 856.27)
MMSE: 21-25	0.64	0.011			Beta(1137.19, 639.67)
MMSE: 26-30	0.69	0.023			Beta(282.51, 126.92)
Institutionalized (MMSE 0-9)	0.33	0.044	Jonsson et al (2006)[201]	Analysis of IPD from Wolstenholme et al suggests participants had an average MMSE of 9 when 0.04 years prior to institutionalization, therefore institutionalization used as proxy for severe Alzheimer's disease	Beta(36.59, 74.28)
Dead	0				
Monthly drug costs					
Memantine (20mg) % 15mg	£75		BNF58	Author judgement on % RCT participants having mean dose of 15mg or 20 mg	%15mg=Normal(0.2,0.05)

Parameter	Value	Std err	Source	Justification	Distribution for PSA*
% 20mg					%20mg = 1 - %15mg
6-monthly monitoring out-patient visit cost	£158		National Schedule of Reference Costs (2008-9).	NHS Trusts consultant led; follow-up attendance non-admitted face-to-face	Gamma(4.94, 32)
Monthly monitoring out-patient visit cost	£26		Calculated from 6-monthly cost of £158 for single visit		
Pre-institutionalized NHS/PSS costs	Cubic eqn with time		Relationship between NHS & PSS costs and time prior to institutionalisation calculated from IPD of Wolstenholme et al[181]		Cholesky matrix shown in Section *7.3.10.2*
Institutionalized	£2,941	£108	Calculated from IPD of Wolstenholme et al[181]	Monthly cost of institutional care. Only 72% of these costs are assumed to be NHS/PSS	Normal(2941, 108)
% institutionalized costs funded by NHS/PSS	0.72		Dementia UK report (as cited by the 2007 NAO report[5;6])	Assumes that 28% of institutional care costs are not funded by the NHS/PSS	Beta(15, 5.83)

* See Appendix 20 for graphical presentation of distributions

7.4. Results

The full results for the cost–utility analysis of AChEIs are presented first, followed by the full results for the cost–utility of memantine. Due to the many assumptions associated with the parameter estimates in the PenTAG model, it is important to be fully aware of the full uncertainty in the model. Because of this, the first set of analyses presented in this section are those from the probabilistic sensitivity analyses of the base-case parameter values for cholinesterase inhibitors in people with mild to moderate Alzheimer's disease and memantine in people with moderate to severe Alzheimer's disease. These results are followed by the deterministic base-case results which are compared to the corresponding mean estimates from the PSA. The deterministic analyses should not be considered to be the primary results of our analyses.

There is also a great deal of structural uncertainty in the PenTAG models which cannot be accounted for in the PSA. Deterministic sensitivity analyses have been undertaken to explore some of the structural and further parameter uncertainty.

7.4.1. Mild to moderate Alzheimer's disease: cholinesterase inhibitors (Decision problem 1a)

7.4.1.1. Probabilistic sensitivity analysis

The cost-effectiveness results of 10,000 simulations for the base-case analysis of the cost–utility of AChEIs in people with mild to moderate Alzheimer's disease are presented in *Figure 72,* showing that there exists a great deal of uncertainty. The cost-effectiveness acceptability curve shown in

FIGURE 73 demonstrates that there is a very low probability that a particular AChEI is the most cost-effective technology, regardless of the threshold willingness to pay. Moreover, at a willingness to pay of £30,000 per QALY gained there is only a 57% probability that best supportive care is the best treatment option, thus indicating that there is a 43% probability that it is *not* the most cost-effective treatment option. At a willingness to pay of £20,000 per QALY gained, best supportive care has a probability of 62% of being the most cost-effective treatment option.

FIGURE 72 *Base-case cost-effectiveness plane for treatment with AChEIs in people with mild to moderate Alzheimer's disease*

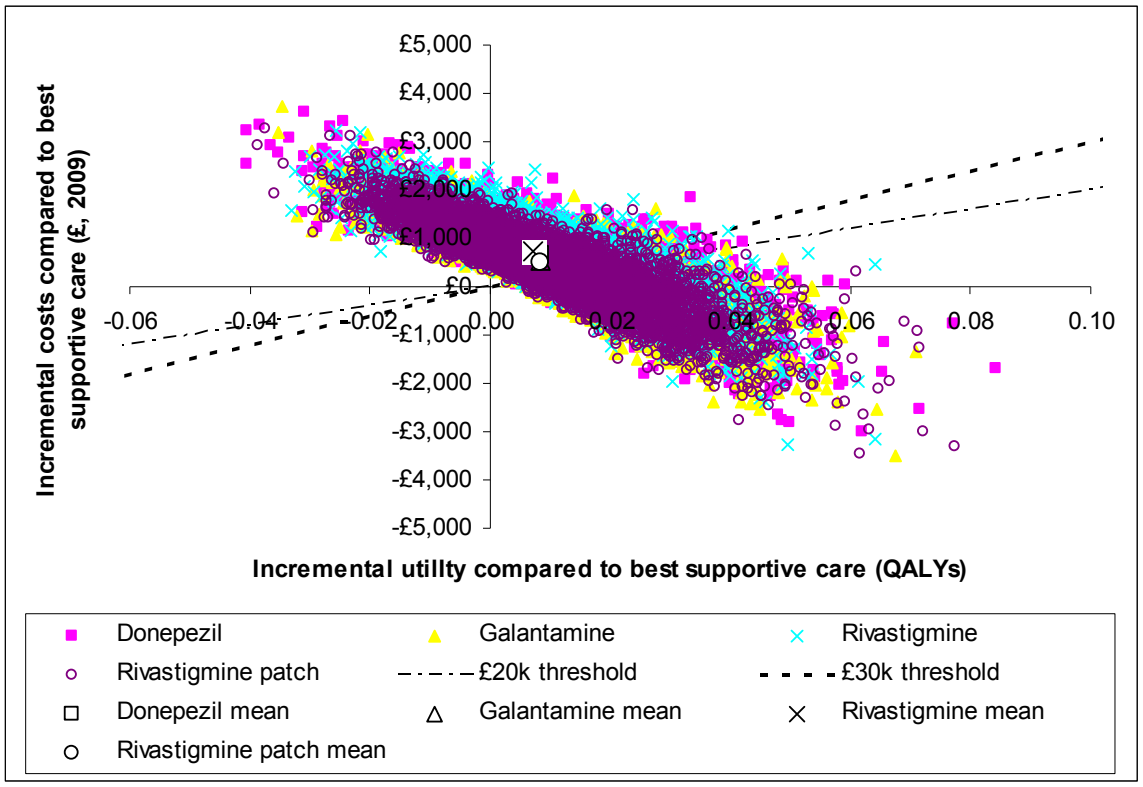
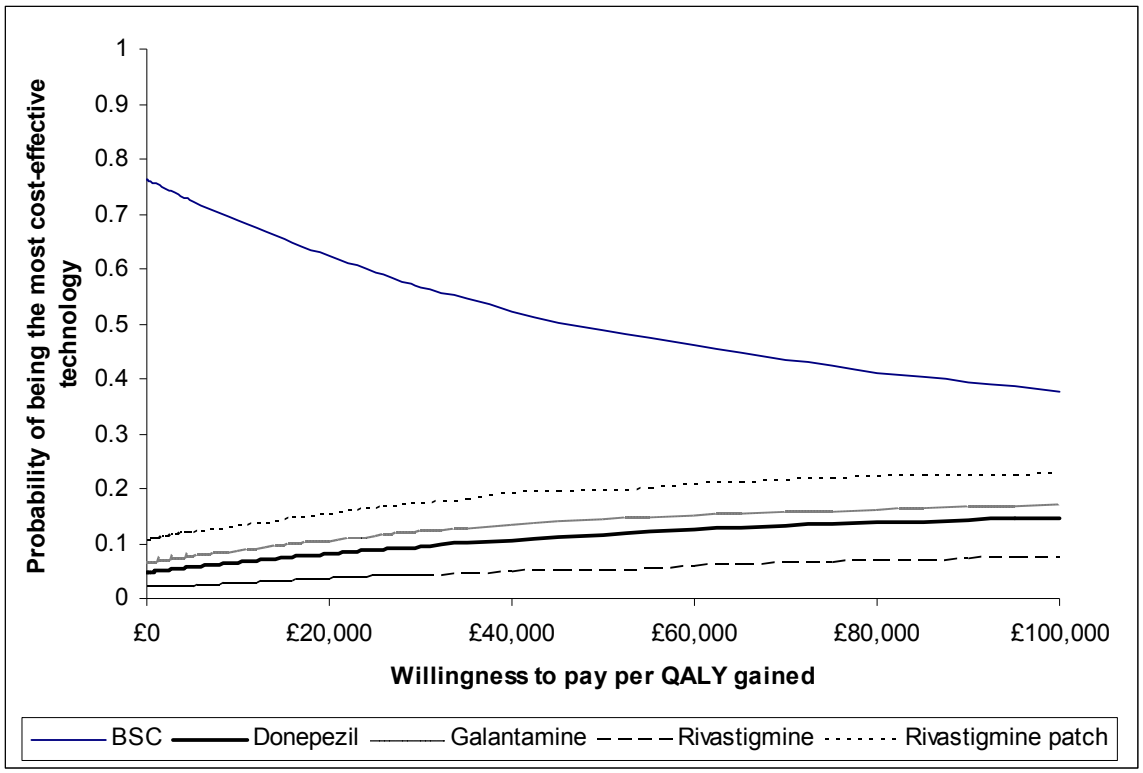


FIGURE 73 *Base-case cost-effectiveness acceptability curve for AChEIs in people with mild to moderate Alzheimer's disease*



BSC, best supportive care

The ICERs obtained from the probabilistic sensitivity analysis (PSA) are £59,800 of rivastigmine patches compared to best supportive care and £157,800 for galantamine compared to rivastigmine patches, where each technology is compared to the next cheapest non-dominated technology. Donepezil and rivastigmine capsules were both estimated to be dominated. As the ICERs obtained from the PSA were not much different to those calculated in the deterministic analysis (see Section *7.4.1.2*), all reference to the base-case analysis will refer to the deterministic ICERs, and not the PSA ICERs.

A cost-effectiveness acceptability frontier is not shown from this PSA as it would follow the line of best supportive care in

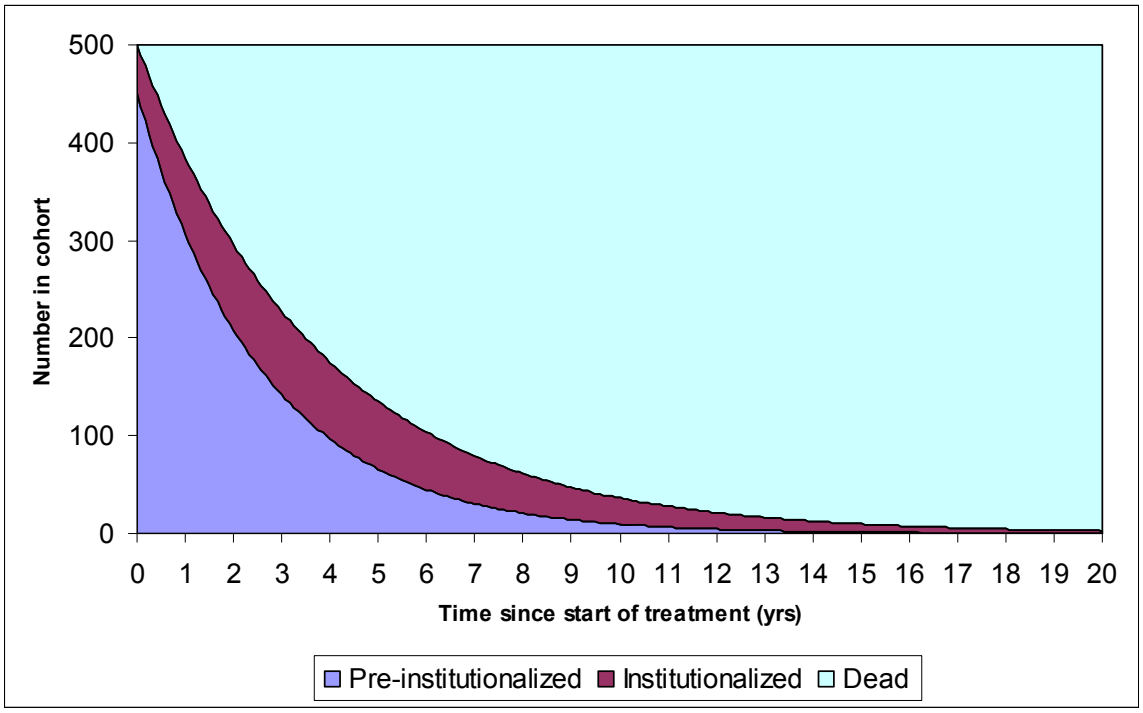
FIGURE 73.

Note that an assumption of perfect correlation between the parameters for the time to institutionalization and the time to death equations had very little impact on the findings from the PSA compared to when no correlation is assumed.

7.4.1.2. Deterministic analysis

A graph of the progression of individuals with mild to moderate Alzheimer's disease from the best supportive care cohort through the three state Markov model is shown in *Figure 74* for the middle age group, having a mean starting age of 77 (representing 50% of the cohort). Ten percent of the cohort start the model in the institutionalized state. Across all three age groups, the mean overall survival for the total prevalent cohort is 3.84 years. This is regardless of the treatment received since, in the base-case analysis, it is assumed that there is no treatment effect on survival. The mean time until the end of pre-institutionalization for the treated cohorts is given in *Table 116*, alongside the total cost and QALY estimates from the deterministic analysis. There is very little difference between the three cholinesterase inhibitors, as might be expected given the similar magnitude of effectiveness for MMSE and ADCS-ADL (refer back to *Table 114*), with treatment leading to a mean of 10-12 days delay in becoming institutionalized.

FIGURE 74 *Progression of the best supportive care cohort for the base-case analysis (mild to moderate Alzheimer's disease, age group 2)*



The base-case results for the incremental cost–utility of the AChEIs compared to the next cheapest, non-dominated technology are given in *Table 116*. Note that the incremental QALYs and the incremental costs are very small for all comparisons, leading to costs per QALY that are particularly sensitive to changes in the parameter inputs (see later sensitivity analyses in Section *7.4.1.4*).

TABLE 116 *Results of the deterministic base-case incremental cost–utility analysis for people with mild to moderate Alzheimer's disease (MMSE 26-10)*

Treatment	Mean months to institutional care[c]	Days delay to institutional care compared to best supportive care	Costs	QALYs	Incremental costs	Incremental QALYs	ICER[ab]
Best supportive care	28.8		£70,237	1.584			
Rivastigmine patch (10cm^2)	29.2	11	£70,717	1.592	£481	0.008	£61,100
Galantamine (16-24mg)	29.2	12	£70,740	1.592	£23	0.0002	£151,100
Donepezil (10mg)	29.2	11	£70,863	1.592	Extended dominated by Rivastigmine patch and Galantamine		
Rivastigmine capsules (9-12mg)	29.1	10	£70,917	1.591	Extended dominated by Rivastigmine patch and Galantamine		

a Cost per QALY rounded to the nearest £100

b Each technology is compared to the next cheapest non-dominated technology

c This compares to a mean time to end of pre-institutionalization for all 92 participants in the study by Wolstenholme and colleagues of 30 months. This difference arises from the Wolstenholme IPD containing all severities of Alzheimer's disease, not just those who are mild to moderate as in the above table.

It is estimated that over a patient's lifetime, treatment with rivastigmine patches for mild to moderate Alzheimer's disease costs £481 more and provides an additional 0.008 QALYs over best supportive care. This leads to an ICER of £61,100. When compared to rivastigmine patches, galantamine provides an additional 0.0002 QALYs, at an extra cost of £23, resulting in galantamine having a very high estimated incremental cost per QALY compared to rivastigmine patches. Donepezil and rivastigmine capsules are both less effective and more costly than rivastigmine patches and galantamine, hence donepezil and rivastigmine capsules are dominated. However, it is important to note that, as the scatter plot in *Figure 75* shows, the costs and benefits of all AChEIs are very similar to the costs and benefits of best supportive care. A 'zoomed-in' version of *Figure 75* is the cost-effectiveness plane of *Figure 76*. This figure demonstrates the cost-effectiveness frontier, yet the scale of the incremental costs and QALYs should be acknowledged: just a difference of 0.008 QALYs and costs less than £500 over a patient's lifetime.

FIGURE 75 *Base-case total costs and QALYs for all treatment options in people with mild to moderate Alzheimer's disease*

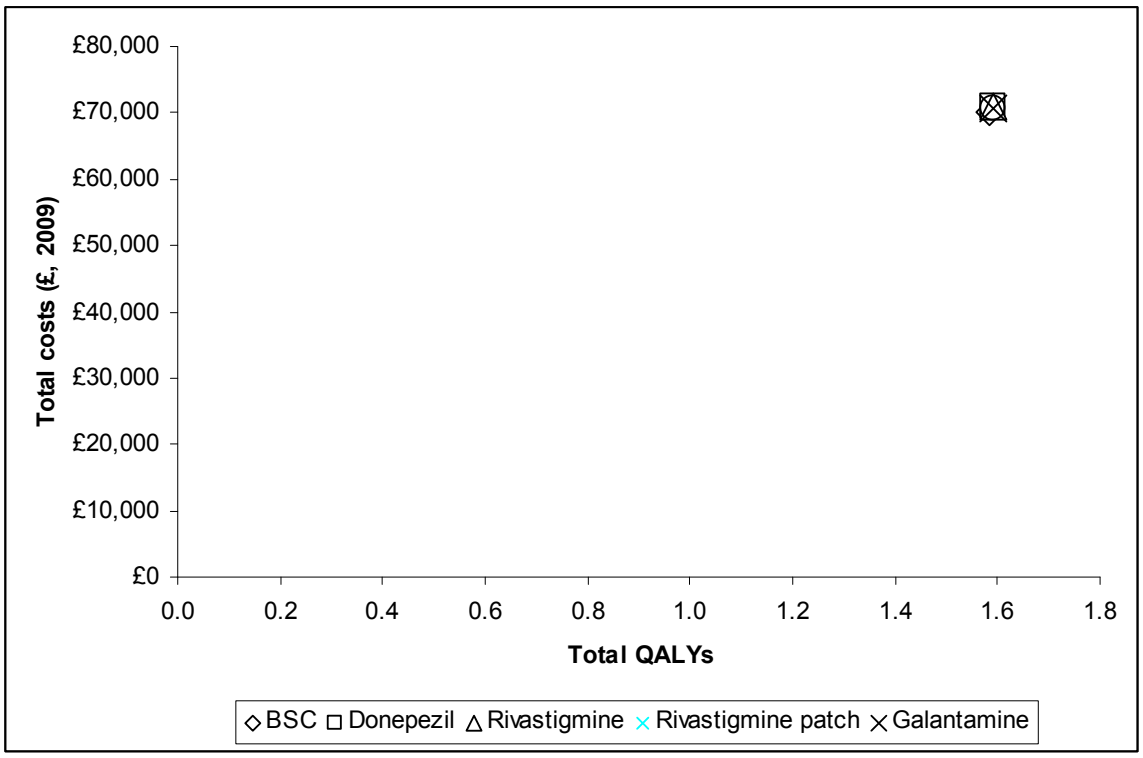
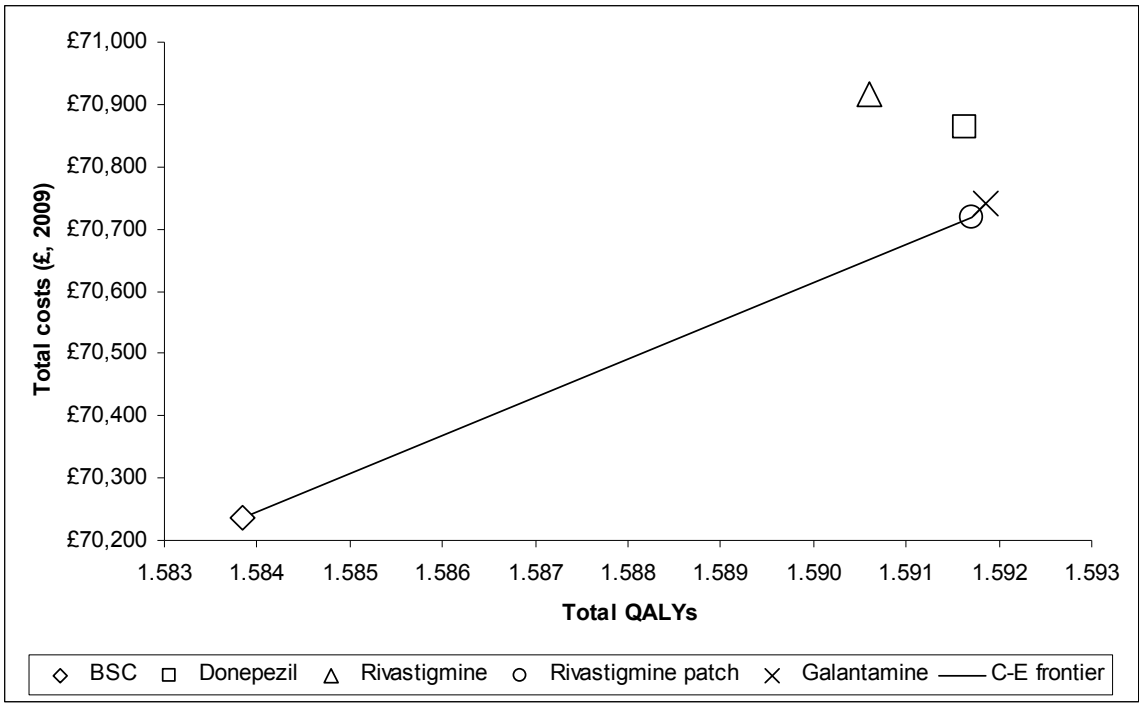


FIGURE 76 *Base-case cost-effectiveness plane for the cost–utility analysis for mild to moderate Alzheimer's disease*



The differences in the component costs between the four AChEIs when each is compared to best supportive care are shown in *Figure 77*. For all four technologies, the largest saving is for the costs associated with being in institutional care. This is as expected since the technologies are estimated to 8delay institutionalisation for 10-12 days. Since overall survival is not assumed to be affected by the AChEIs an individual's total time spent in institutional care is reduced by receiving treatment. The delay to institutionalization is also reflected in the higher costs incurred for the pre-institutionalized state when compared to best supportive care (*Figure 77*). However, the additional costs incurred in the pre-institutionalized state are much lower than those saved from delaying institutionalization, because the cost per unit time is lower in the pre-institutionalized state than in the institutionalized state. Note that institutionalization costs saved for rivastigmine capsules are slightly lower than those for donepezil, galantamine and rivastigmine patches since rivastigmine capsules were estimated to be slightly less effective than the other three technologies (refer back to *Table 114*). *Figure 77* highlights slight difference in drug costs between the AChEIs with rivastigmine capsules being the most expensive and rivastigmine patches the least expensive. The additional QALY gains over best supportive care for the four technologies are all in the pre-institutionalized state (see *Figure 78*). The QALYs lost in

the institutionalized state with treatment with the AChEIs compared to best supportive care reflect the reduced time spent in institutionalization for those on treatment (because the base-case assumptions include no treatment effect on overall survival). Overall, the QALY gains before institutionalization are greater than the QALY losses while in the institutionalized state because the utilities before institutionalization are greater than the utility whilst institutionalized.

FIGURE 77 *Base-case cost components for the cholinesterase inhibitors compared to best supportive care for mild to moderate Alzheimer's disease*

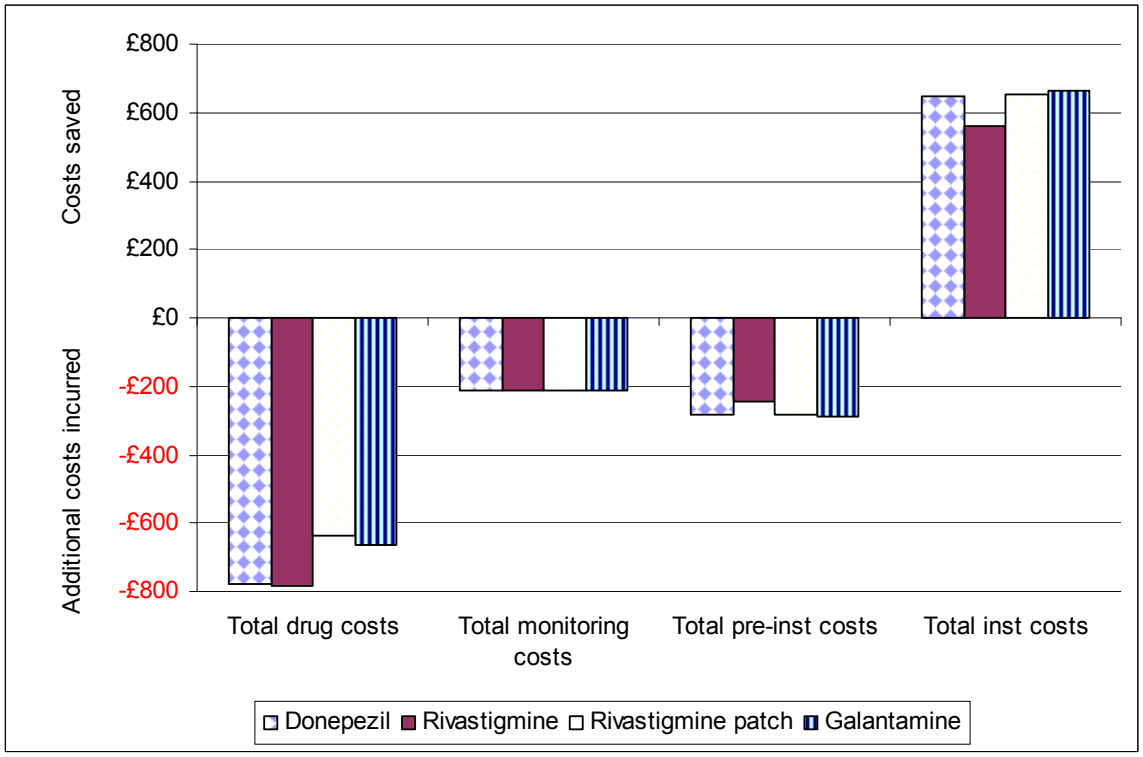
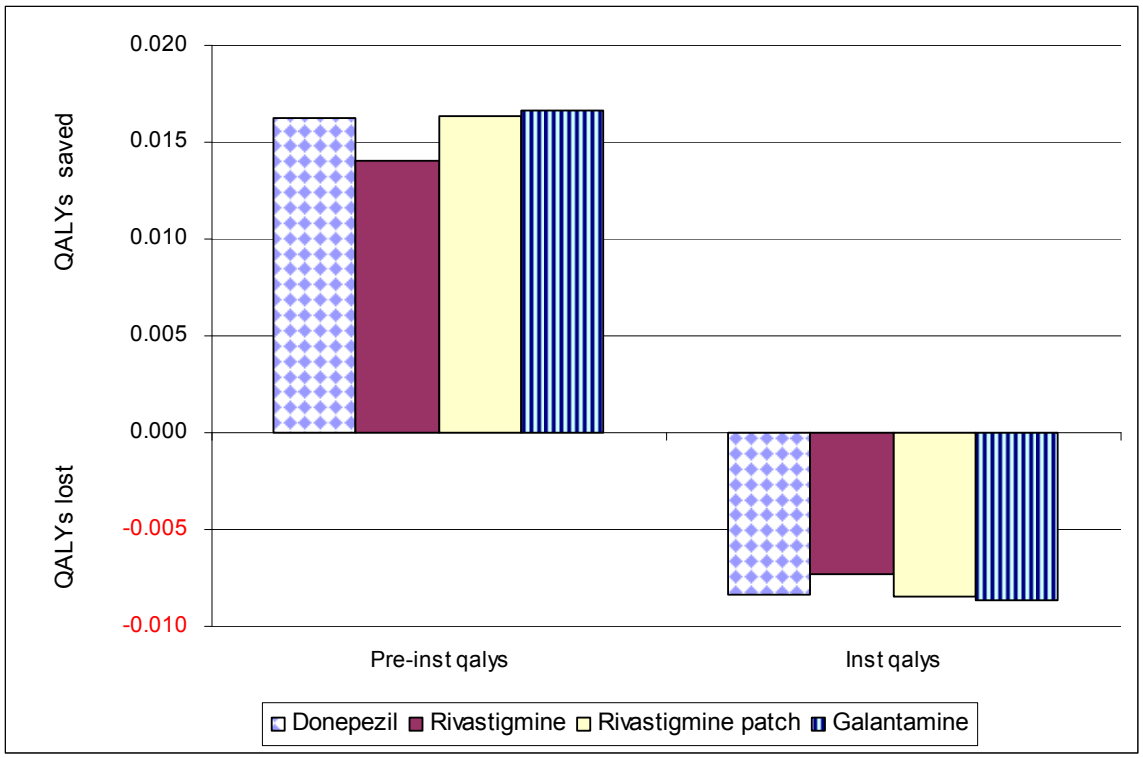


FIGURE 78 *Base-case QALY components for the cholinesterase inhibitors compared to best supportive care in mild to moderate Alzheimer's disease*



7.4.1.3. Summary of deterministic and probabilistic analyses

The probabilistic sensitivity analysis and the deterministic base-case analysis indicates that very little QALY gain is seen when individuals with mild to moderate Alzheimer's disease are treated with AChEIs. There are similarly small incremental costs associated with AChEI treatment. Although treatment does delay time to institutionalization, therefore leading to reduced time spent in the most expensive state (institutionalization), the QALY gains are very small compared to the extra costs of treatment, monitoring and time spent in pre-institutionalized care. Rivastigmine patches are most likely to be considered cost-effective, but this is with an ICER of £61,000 per QALY. However, as the PSA and following sensitivity analyses demonstrate, there is a great deal of uncertainty associated with the above costs per QALYs. In *Table 117* , the deterministic and probabilistic ICERs from the PenTAG model are presented. Note that these estimates are also presented in Section *7.6* for comparison with the SHTAC estimates from the previous review [2] and the industry estimates for donepezil.

TABLE 117 *Base-case ICERs[a] from the PenTAG model for AChEIs in people with mild to moderate Alzheimer's disease*

	PenTAG		
	Deterministic[b]	**Probabilistic[b]**	**Deterministic v. BSC[c]**
Rivastigmine patches (10cm^2)	£61,100	£59,800	£61,100
Galantamine (16–24mg)	£151,100	£157,800	£62,700
Donepezil (10mg)	Dominated	Dominated	£80,400
Rivastigmine capsules (9-12mg)	Dominated	Dominated	£100,600

a Rounded to nearest £100

b Compared to next cheapest, non-dominated treatment option

c BSC=best supportive care

7.4.1.4. One-way deterministic sensitivity analyses

Treatment effect on mortality

In the base-case analysis, it was assumed that there was no treatment effect on survival. However, analysis of the IPD from Wolstenholme and colleagues for predicting time to death used MMSE, Barthel ADL and age as independent variables, and the effectiveness data indicate that AChEI treatment affects MMSE and ADL. Thus, as a sensitivity analysis it is assumed that treatment effect measured by MMSE and the Barthel ADL does affect survival. The mean times to institutionalization do not change from the base-case analysis (see *Table 116*), but the mean time to death is extended and given in *Table 118* for each treatment cohort. All treatments delay death (by 22-26 days) compared to best supportive care.

The cost–utility analysis results assuming a treatment effect on survival are shown in *Table 118*.

TABLE 118 *Incremental cost–utility analysis for mild to moderate disease when survival effect of treatment is assumed*

Treatment	Mean time to death[c]	Extended life compared to BSC[d]	Costs	QALYs	Incremental costs	Incremental QALYs	ICER[ab]
Best supportive care	46.0		£70,237	1.584			
Rivastigmine patch (10cm^2)	46.9	26	£72,130	1.610	£1,893	0.026	£72,200
Galantamine (16-24mg)	46.9	26	£72,176	1.611	£46	0.0005	£101,600
Rivastigmine (9-12mg)	46.7	22	£72,100	1.606	Dominated by rivastigmine patch and galantamine		
Donepezil (10mg)	46.8	25	£72,200	1.609	Extended dominated by rivastigmine patch and galantamine		

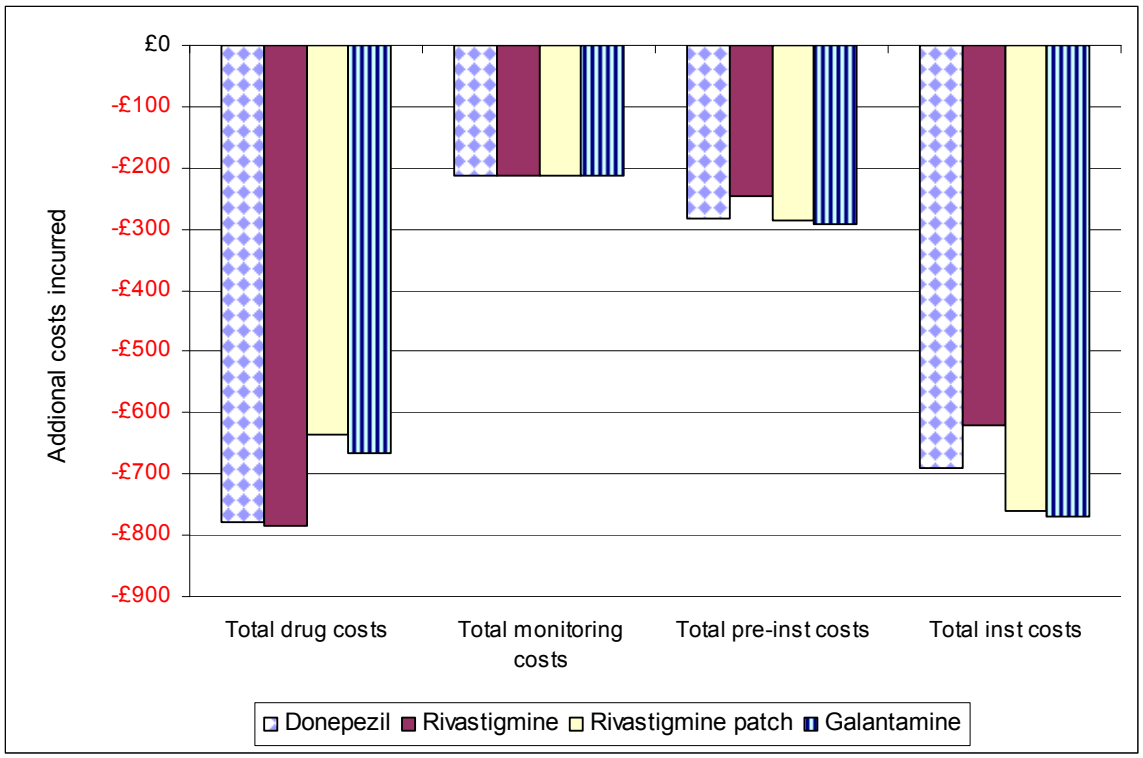
a Cost per QALY rounded to the nearest £100

b Each technology is compared to the next cheapest non-dominated technology

c This compares to a mean time to end of pre-institutionalization for all 92 participants in the study by Wolstenholme and colleagues of 30 months. This difference arises from the Wolstenholme IPD containing all severities of Alzheimer's disease, not just those who are mild to moderate as in the above table.

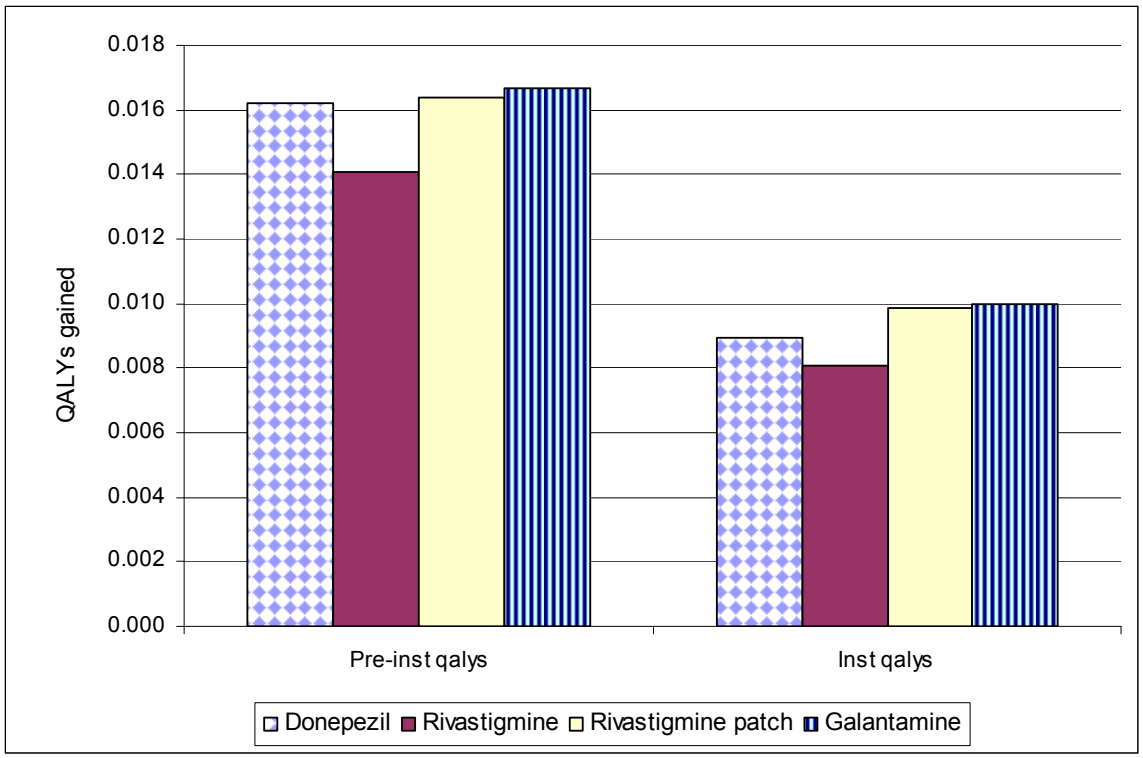
d BSC = best supportive care

FIGURE 79 *Cost components for the cholinesterase inhibitors compared to best supportive when a treatment effect on survival is assumed*



It is estimated that treatment with rivastigmine patches provides an additional 0.026 QALYs over a patient's lifetime compared to best supportive care at an additional cost of £1,893, resulting in an incremental cost per QALY of £72,200. Galantamine provides an additional 0.0005 QALYs over rivastigmine patches, but at an added cost of £46, giving an ICER of £101,600. As in the base-case analysis donepezil and rivastigmine capsules are either dominated or extended dominated by rivastigmine patches and galantamine. In comparison to the base-case analysis, more QALYs are gained when a treatment effect on survival is assumed, due to additional life, but this gain is spent in a more expensive state, institutional care (see *Figure 79* and *Figure 80*).

FIGURE 80 *QALY components for the cholinesterase inhibitors compared to best supportive care assuming a treatment effect on survival*



Cognitive effectiveness based on the ADAS-cog outcome

The cost-effectiveness results for the AChEIs are shown in *Table 119* where the treatment effect on cognition is measured using ADAS-cog rather than MMSE (as in the base-case analysis). It can be seen that galantamine dominates the other three AChEI technologies. This contrasts with the base-case analysis where MMSE was used, since galantamine is estimated to be the most effective AChEI when measured by ADAS-cog (see *Table 108*). However, the ICER for galantamine compared to best supportive care is still above the £30,000 per QALY threshold, £58,400 per QALY, even though it dominates the other treatment alternatives.

TABLE 119 *Incremental cost–utility analysis for mild to moderate disease when effectiveness on cognition is measured by the ADAS-cog*

Treatment	Costs	QALYs	Incremental costs	Incremental QALYs	ICER[ab]
Best supportive care	£70,237	1.584			
Galantamine (24mg)	£70,724	1.592	£488	0.008	£58,400
Rivastigmine patch (10cm^2)	£70,767	1.591	Dominated by galantamine		
Rivastigmine (≤12mg)	£70,923	1.590	Dominated by galantamine		
Donepezil (10mg)	£70,867	1.592	Dominated by galantamine		

a Cost per QALY rounded to the nearest £100

b Each technology is compared to the next cheapest non-dominated technology

The patterns for the cost and utility components for this sensitivity analysis (see *Figure 81* and *Figure 82*) are similar to those for the base-case analysis: costs are saved in institutional care, and QALYs are gained in the pre-institutionalization state due to delay to institutionalization, but lost in the institutionalized state due to having a reduced time in this state.

FIGURE 81 *Cost components of the cholinesterase inhibitors compared to best supportive care when ADAS-cog is used to measure treatment effect on cognition*

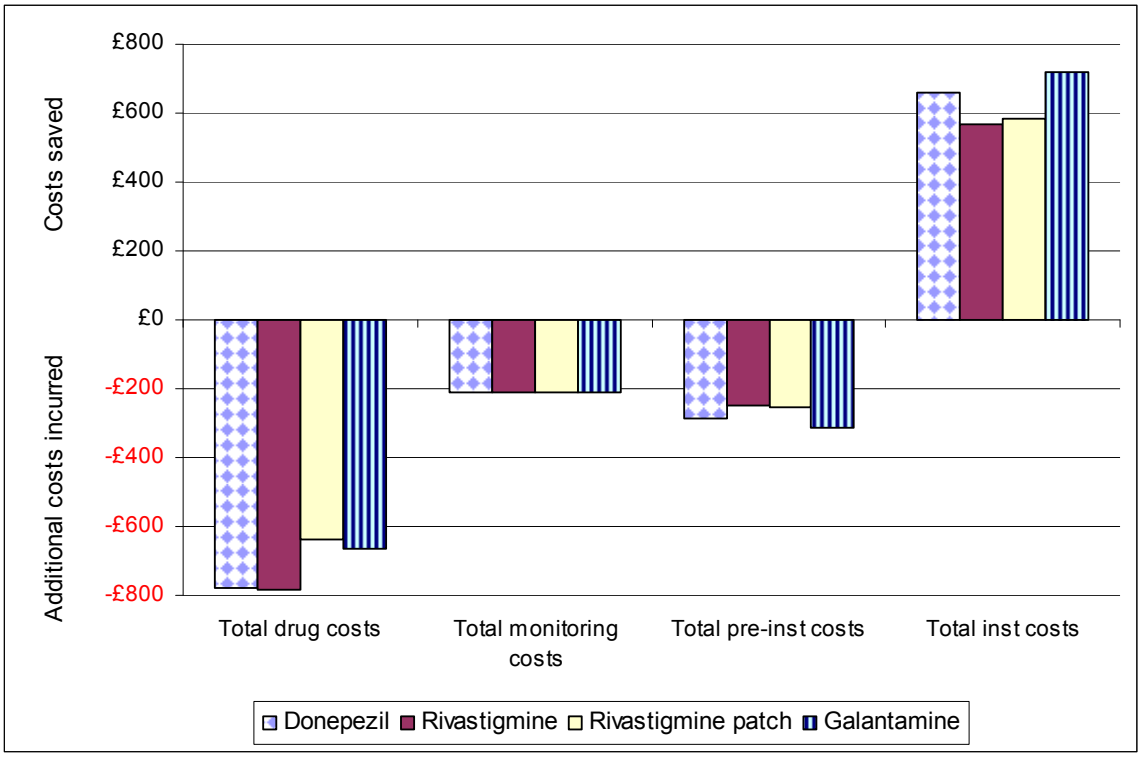
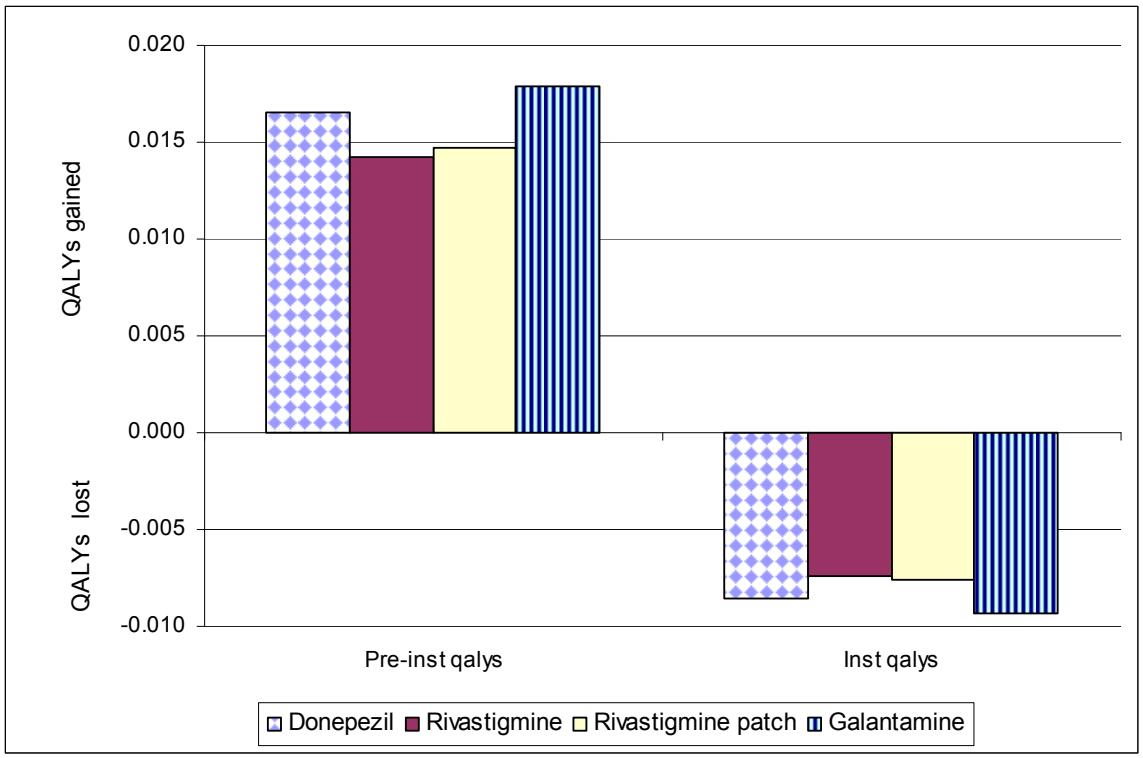


FIGURE 82 *QALY component of the cholinesterase inhibitors compared to best supportive care when ADAS-cog is used to measure treatment effect on cognition*



Further one-way sensitivity analyses

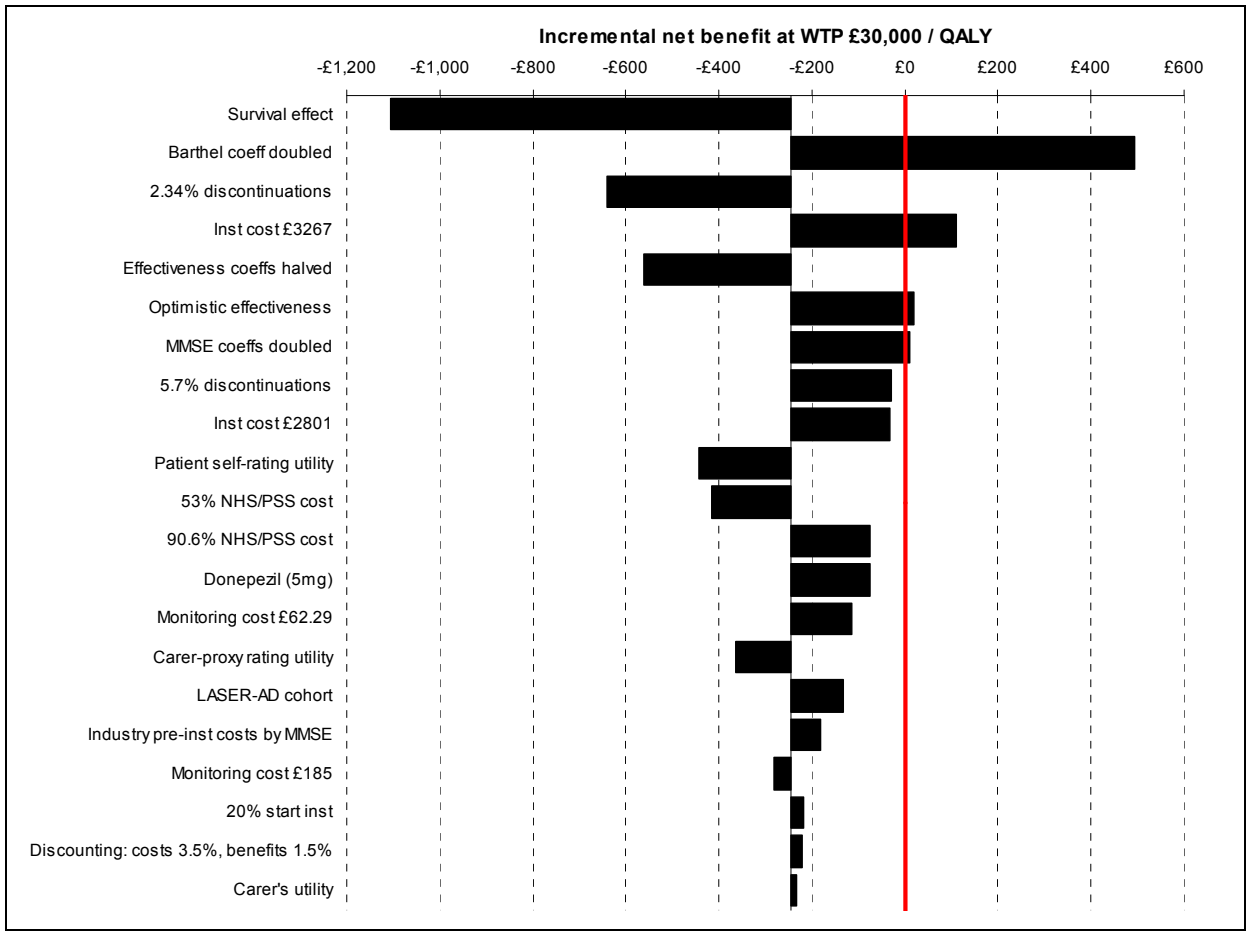
The parameter values and assumptions explored in the following one-way sensitivity analyses are shown in *Table 120*. Analyses are presented as incremental net monetary benefits at a willingness to pay of £30,000 per QALY, with each technology compared to the next cheapest non-dominated technology (as in the above analyses). In the majority of one-way sensitivity analyses, rivastigmine patches were compared to best supportive care and galantamine compared to rivastigmine patches. As with the base-case analysis and the above sensitivity analyses, rivastigmine and donepezil were often dominated in the scenarios reported below. Tornado plots are presented for rivastigmine patches compared to best supportive care (*Figure 83*) and galantamine compared to rivastigmine patches (*Figure 84*). Tornado plots for all AChEIs compared to best supportive care are in Appendix 21.

TABLE 120 *Parameter and assumption changes for deterministic sensitivity analyses for base-case analysis of AChEIs for mild to moderate Alzheimer's disease*

Parameter/ assumption	Base-case	Deterministic sensitivity analysis	Reference in tornado plots
Survival effect	Independent of treatment	Survival within first six months based on baseline characteristics regardless of treatment. From six months survival based on baseline characteristics PLUS treatment effect	Survival effect
Basis for cognitive effectiveness	Parameterised on MMSE	Cognitive effectiveness parameterised on ADAS-cog	Effectiveness on ADAS-cog
Drug costs	See *Table 113*	Industry cost for donepezil; 9mg cost for rivastigmine capsules; 16mg cost for galantamine; 24mg cost for galantamine	Drug cost
Cost in institutional care	£2,941 per month estimated from Wolstenholme IPD	£3267 from Lundbeck submission; £2801 from Eisai submission	Inst cost £
% institutional costs NHS/PSS	0.72	0.53 & 0.906	% NHS/PSS cost
Coefficients for MMSE and Barthel-ADL in the prediction of time to institutionalization	Mean estimates from statistical analyses	MMSE and/or Barthel coefficients doubled or halved	Effectiveness coeffs halved/doubled; Barthel coeff halved; MMSE coeff halved
Treatment discontinuations	4% of the total cohort per month	The maximum (5.7%) and minimum (2.34%) from the RCTs discontinue each month	% discontinuations
Effectiveness	As summarised from the RCTs	Optimistic effectiveness: MMSE effectiveness + 0.5, Barthel effectiveness + 1; Pessimistic effectiveness: MMSE effectiveness – 0.5, Barthel effectiveness - 1	Optimistic effectiveness; Pessimistic effectiveness
Cost in pre-institution state	Based on relationship from Wolstenholme IPD	Transformed industry pre-inst costs by MMSE to time to inst	Industry pre-inst costs by MMSE
Population characteristics	Based on Wolstenholme IPD	Based on LASER-AD IPD	LASER-AD cohort
Severity of cohort	Mild to moderate	Mild or moderate	Cohort severity
% cohort start in institutional care	10%	20%	% start inst
Monitoring costs	From National Schedule Reference Costs,	£185, upper value of interquartile range from National Schedule	Monitoring cost £

Parameter/ assumption	Base-case	Deterministic sensitivity analysis	Reference in tornado plots
	£158 per visit	Reference Costs; £62.29, Lundbeck estimate of subsequent out-patient visit	
Patient utility weights	EQ-5D carer-proxy utilities	Patient self-rated EQ-5D utility and alternative sources of carer-proxy utility (AQoL)	Patient self-rated utility; Carer-proxy utility
Carer utility weights	Not included	HUI:2	Carer utility

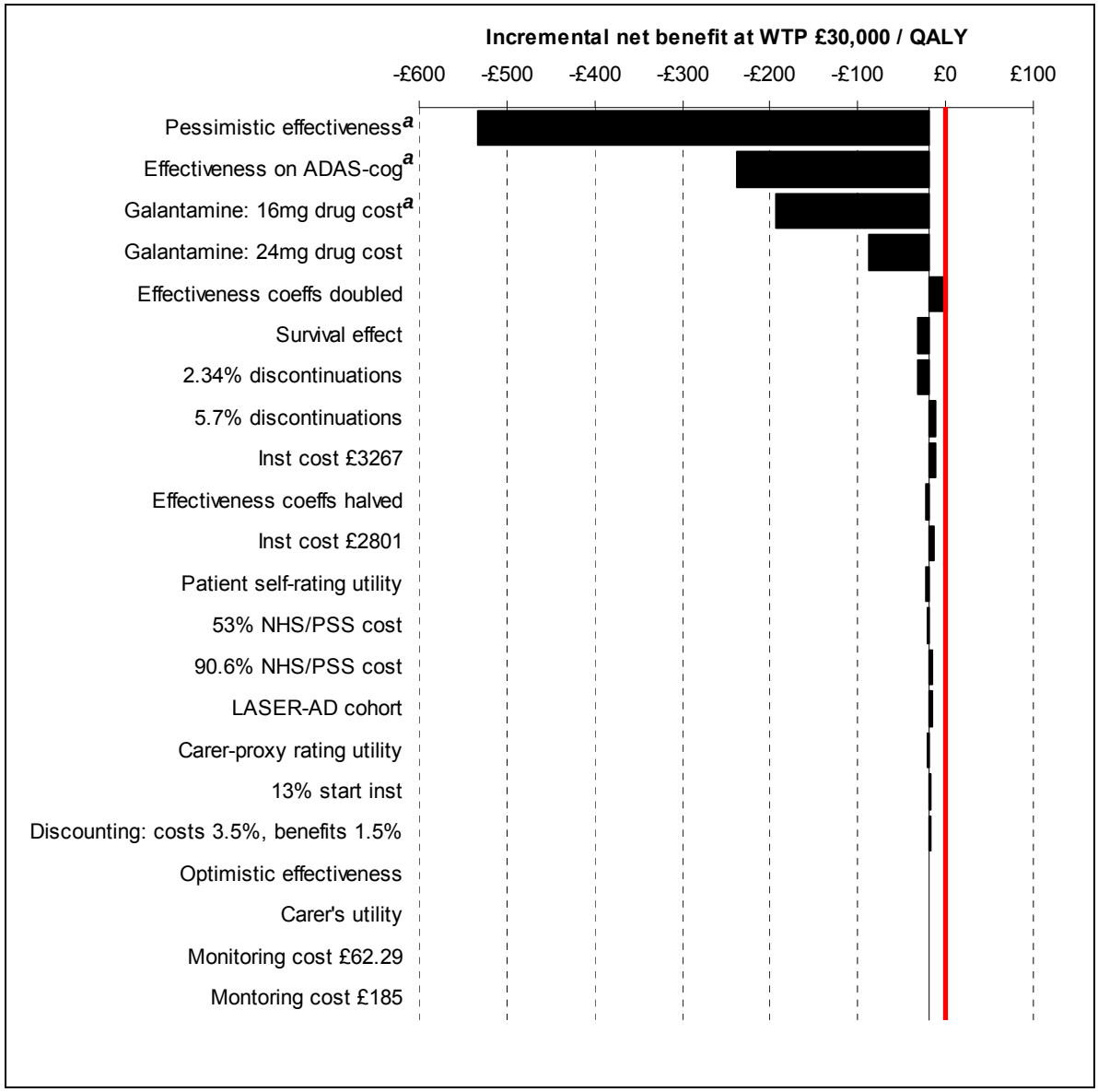
FIGURE 83 *One-way sensitivity analyses for the incremental net monetary benefit of rivastigmine patches compared to best supportive care for mild to moderate Alzheimer's disease*



See *Table 120* for a description of the individual sensitivity analyses undertaken.

Note that there is no bar for 'Pessimistic effectiveness' or 'Effectiveness on ADAS-cog' in **FIGURE 83**, as galantamine dominates rivastigmine patches in both sensitivity analyses.

FIGURE 84 *One-way sensitivity analyses for the incremental net monetary benefit of galantamine compared to rivastigmine patches*



[a] Compared to best supportive care, the next cheapest non-dominated treatment option

See Table 120 for explanation of description of individual sensitivity analyses undertaken

At a willingness to pay of £30,000 per QALY gained, rivastigmine patches have an incremental net benefit of -£224 compared to best supportive care. Galantamine has an incremental net benefit of -£18 compared to rivastigmine patches at a willingness to pay of £30,000 per QALY. Note that for the three assumptions having the largest impact on the net benefit for galantamine, galantamine is compared with best supportive care, as galantamine dominates rivastigmine patches. Given the similar costs and QALYs associated with

galantamine and rivastigmine patches, it is not surprising that in some situations galantamine may dominate rivastigmine patches.

As shown above, the assumption of a treatment effect on survival leads to the AChEIs having a larger cost per QALY gained than in the base-case analysis. *Figure 83* and *Figure 84* demonstrate that of all the scenarios assessed in the one-way sensitivity analyses; this assumption has the largest impact on the net benefit. Also of great importance in the comparison of rivastigmine patches to best supportive care is the Barthel ADL coefficient in the equations predicting time to institutionalization. The larger this coefficient, the more of an impact the transformed ADCS-ADL treatment effect has in predicting time to institutionalization. It therefore follows that assuming a greater impact of the ADL treatment effect leads to a larger net benefit for the AChEIs compared to best supportive care.

The percentage of the total cohort discontinuing treatment each month is a very important factor when AChEIs are compared to best supportive care (*Figure 83* and Appendix 21**)**. As pointed out in Section *7.3.7.2*, this parameter only affects the costs associated with treatment, not the effectiveness, since it is assumed that the effect estimates are based on an ITT analysis. Therefore, lower estimates of this percentage lead to greater treatment and monitoring costs, resulting in a negative net benefit for the AChEIs. Higher estimates lead to fewer costs and greater net benefit associated with the AChEIs.

Assumptions on the costs of care in the institutionalized state have a large impact on the results as would be expected. Since it is assumed that the AChEIs delay and therefore, in the base-case analysis, reduce time spent in institutionalized care, this cost is important. Assuming a greater cost for institutional care compared to pre-institutional care leads to more costs saved by the treatments. This is demonstrated in *Figure 83* where higher costs in institutionalized care (either by assuming a greater total cost or by increasing the percentage of institutionalized costs funded by NHS/PSS) lead to greater net benefit at a willingness to pay of £30,000 per QALY gained.

This is also some uncertainty as to the utility estimates used. Alternative estimates of carer-proxy utility also led to lower estimates of net benefit since these estimates provide less of a change in utilities as the disease progresses. Therefore, buy delaying disease progression, a greater utility gain is obtained when there is a larger difference between utility for mild disease compared to severe disease. The estimates used in the base-case analysis span a large range of utility weights across severity, from 0.69 for MMSE>25 to 0.33 for MMSE<10.

These utility estimates are therefore more favourable to the AChEIs in the PenTAG model since a delay to more severe stages of AD leads to a bigger gain in utility than would be obtained using alternative care-proxy estimates having a narrower range of values across severity. Use of patient's self-rated quality of life lead to lower estimates of net benefit for the AChEIs since even for the most severe state a utility of 0.78 was reported compared to 0.84 for MMSE of 26-30. Thus, fewer QALY gains by delaying entry into institutional care are obtained when assuming patient rated quality of life estimates. This is not a surprising result since, the most severe state was estimated to have greater utility by patients than the adjacent less severe state (refer back to Section *7.3.9.1*).

Inclusion of carer's own quality of life estimates led to a very small increase in the net benefit of the AChEIs. This is as expected given that these estimates are based on data indicating that there is very little change in carer's quality of life as the disease progresses.

When AChEIs are compared to other AChEIs many of the assumptions important in comparison with best supportive care are no longer important (see *Figure 84*).

7.4.1.5. Summary of one-way sensitivity analyses

In *Table 121* the degree of uncertainty in the decision model and the impact of these parameters on the cost-effectiveness of the AChE inhibitors is presented for people with mild to moderate Alzheimer's disease. The most important items are those discussed above, the main one being whether a treatment effect on survival is assumed. This structural uncertainty is not accounted for in the PSA, and so the PSA should be considered to describe the minimum amount of overall uncertainty in the PenTAG model.

TABLE 121 *Degree of uncertainty in model assumptions and impact on the cost-effectiveness of the AChEIs*

Issue	Evidence source	Level of uncertainty in data	Impact of uncertainty on cost-effectiveness	Overall rating of importance in cost-effectiveness
Assuming a treatment effect on survival	No published RCT or epidemiological evidence. Survival prediction allows treatment survival effect	High	High	Very important
Treatment discontinuations	Final time-point data from RCTs	High	High	Very important
Costs in institutional care	Inflated 20-year old estimates from 92 individuals	High	High	Very important
Prediction of time to institutionalization based on MMSE and Barthel ADL	Statistical analysis of 92 individuals	High	High	Very important
Effectiveness evidence	Mix of different quality RCTs	Moderate	High	Important
Patient's health state utility	Proxy respondents or self-rated from published literature	Moderate	Moderate	Important
Carer's health state utility	Poor published evidence	High	Low	Moderate importance
% of costs in institutional care funded by NHS/PSS	Poor published evidence plus expert opinion	High	High	Important
Costs in pre-institutional state	Inflated 11- to 20-year old estimates from 92 individuals	High	Moderate	Important
Cost of treatment monitoring visit	National Schedule Reference Costs	Low	Moderate	Moderately important
% starting model in institutional care	Published epidemiological study and author assumption	High	Low	Not important
Baseline characteristics	Statistical analysis of 92 individuals	Low	Low	Not important
Cost of drugs	BNF compared to some poor reporting of doses used in RCTs	Moderate	Moderate	Moderately important

7.4.2. Moderate to severe Alzheimer's disease: memantine (Decision problem 2a)

7.4.2.1. Probabilistic sensitivity analysis

As with the cost–utility analysis of AChEIs there is a great deal of uncertainty associated with estimation of the costs and QALYs of treatment with memantine compared to best supportive for people with moderate to severe Alzheimer's disease (see *Figure 85*). The cost-

effectiveness acceptability curve (*Figure 86*) indicates that there is very little evidence to suggest that memantine would be the most cost-effective option when compared to best supportive care, regardless of the willingness to pay per QALY gained. There is <4% probability that memantine is the most cost-effective treatment at a willingness to pay threshold of £30,000 per QALY. At a willingness to pay of £20,000 per QALY gained, memantine has a probability of 2.6% of being the most cost-effective treatment option.

FIGURE 85 *Base-case cost-effectiveness plane for memantine in people with moderate to severe Alzheimer's disease*

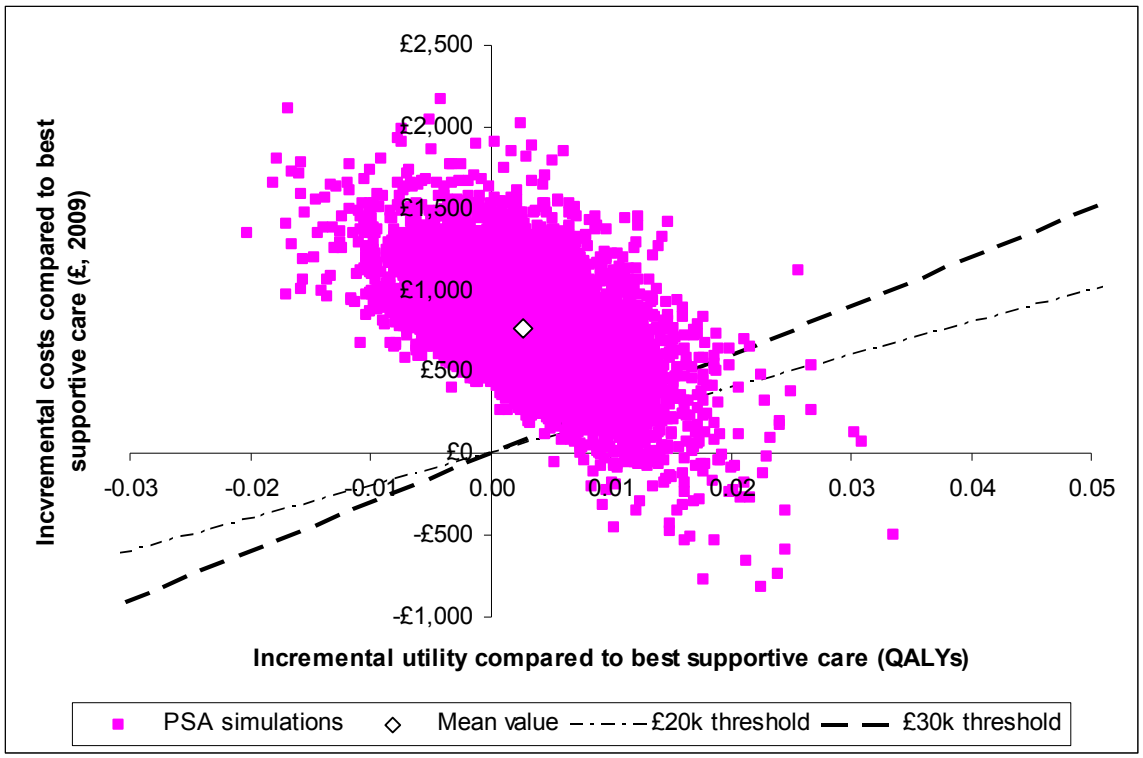
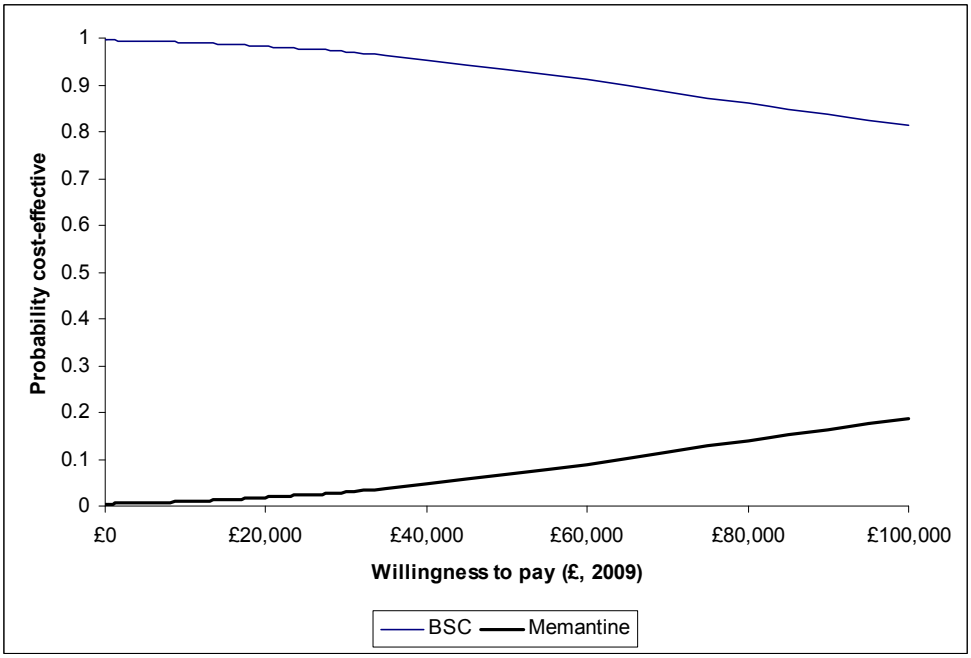


FIGURE 86 *Base-case cost-effectiveness acceptability curve for memantine in people with moderate to severe Alzheimer's disease*



The ICER from the PSA is estimated to be £288,800, which is similar to the estimate from the deterministic base-case analysis (see Section *7.4.2.2*). Note again that the cost-effectiveness acceptability frontier is not presented as it would resemble the line for best supportive care in *Figure 86*.

7.4.2.2. Deterministic analysis

The progression of a proportion of the moderate to severe cohort on best supportive care through the model is represented graphically in *Figure 87* as an example of the time spent within each state of the model. *Figure 87* is based on data for individuals with a mean starting age of 78 (representing 50% of the cohort). Forty percent of the cohort are assumed to be in institutional care at the start of the model. The mean overall survival across all three age cohorts for moderate to severe Alzheimer's disease is 42.1 months. The mean time to institutionalization for the best supportive care cohort is 17.7 months, while for the memantine cohort this is 17.8 months, a delay to institutionalization of about 6 days.

FIGURE 87 *Progression of the best supportive care cohort in the base-case (moderate to severe Alzheimer's disease, age group 2)*

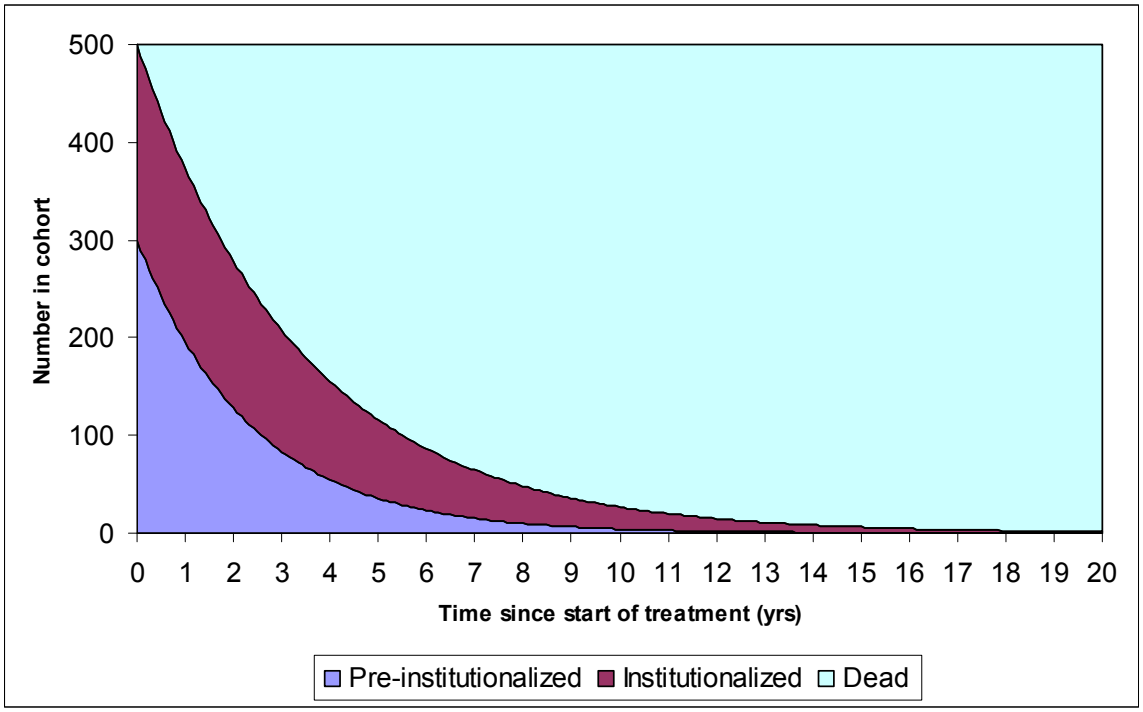
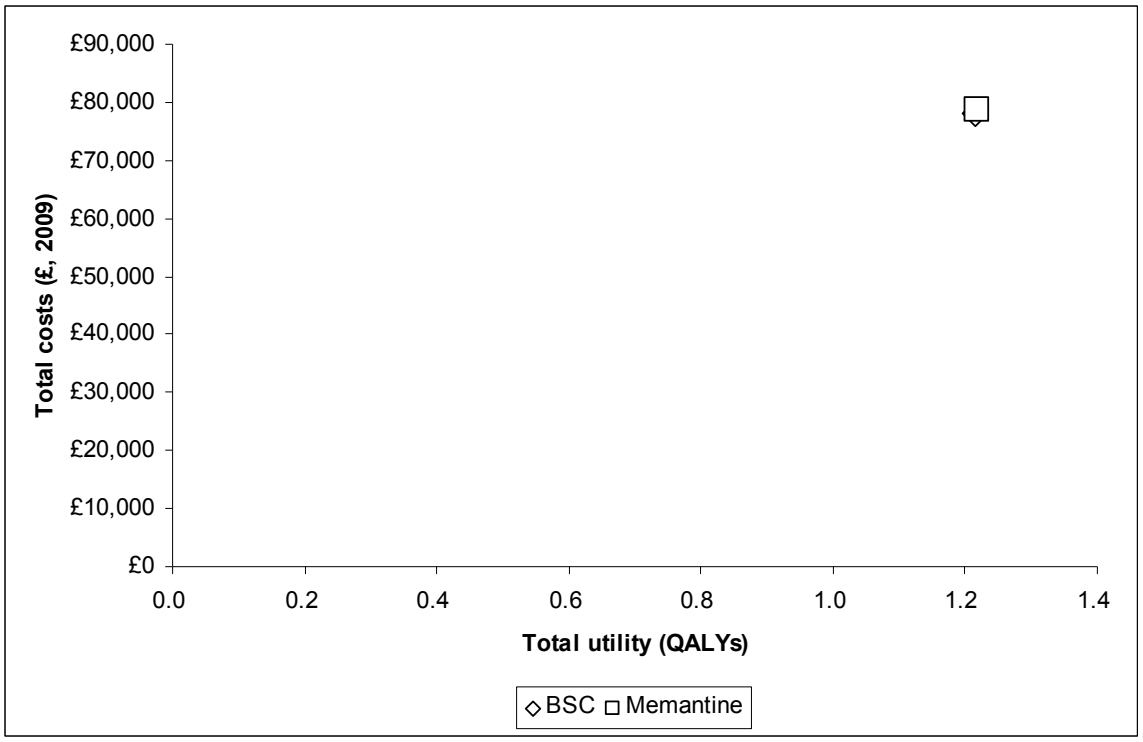


TABLE 122 *Results of the base-case deterministic analysis for people with moderate to severe Alzheimer's disease (MMSE 20-0)*

Treatment	Costs	QALYs	Incremental costs	Incremental QALYs	ICER[a]
Best supportive care	£78,136	1.214			
Memantine (20mg)	£78,855	1.217	£719	0.003	£248,500

a Cost per QALY rounded to the nearest £100

FIGURE 88 *Base-case cost-effectiveness plane for moderate to severe Alzheimer's disease*



The base-case cost–utility analysis result for memantine compared to best supportive care for people with moderate to severe Alzheimer's disease (MMSE 20-0) is given in *Table 122* and shown in *Figure 88*. For a gain of 0.003 QALYs over a patient's lifetime when treated with memantine compared to best supportive care, the extra cost is £719, leading to an estimated cost per QALY of £248,500 from the deterministic base-case analysis. As with the base-case analysis results for AChEIs, the incremental QALYs gained and the additional costs associated with memantine are very small. The cost components detailed in *Figure 89* demonstrate that, as with the AChEIs, the cost savings of treatment with memantine occur while the individual is in institutionalized care. However, the drug, monitoring and pre-institutionalized costs reduce these savings. The gains in QALYs with memantine over best

supportive care (see *Figure 90*) are seen in the pre-institutionalised state, since longer time is spent in this state for memantine-treated individuals.

FIGURE 89 *Base-case cost components for memantine compared to best supportive care for moderate to severe Alzheimer's disease*

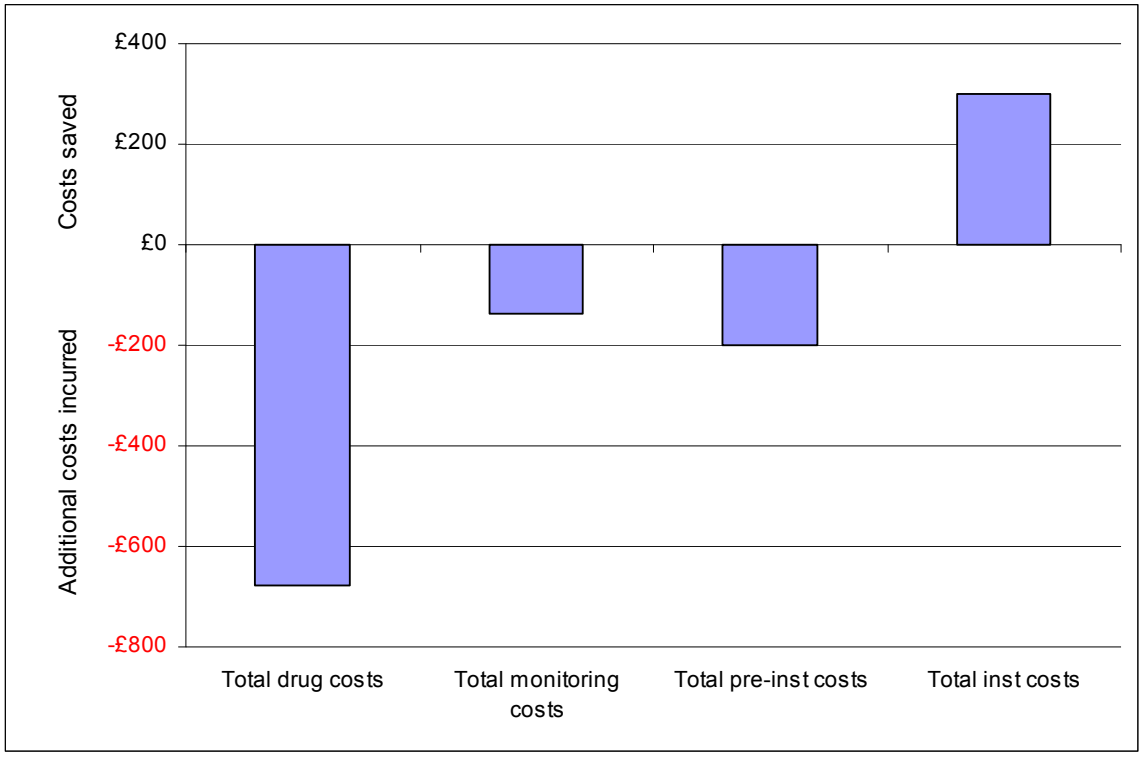
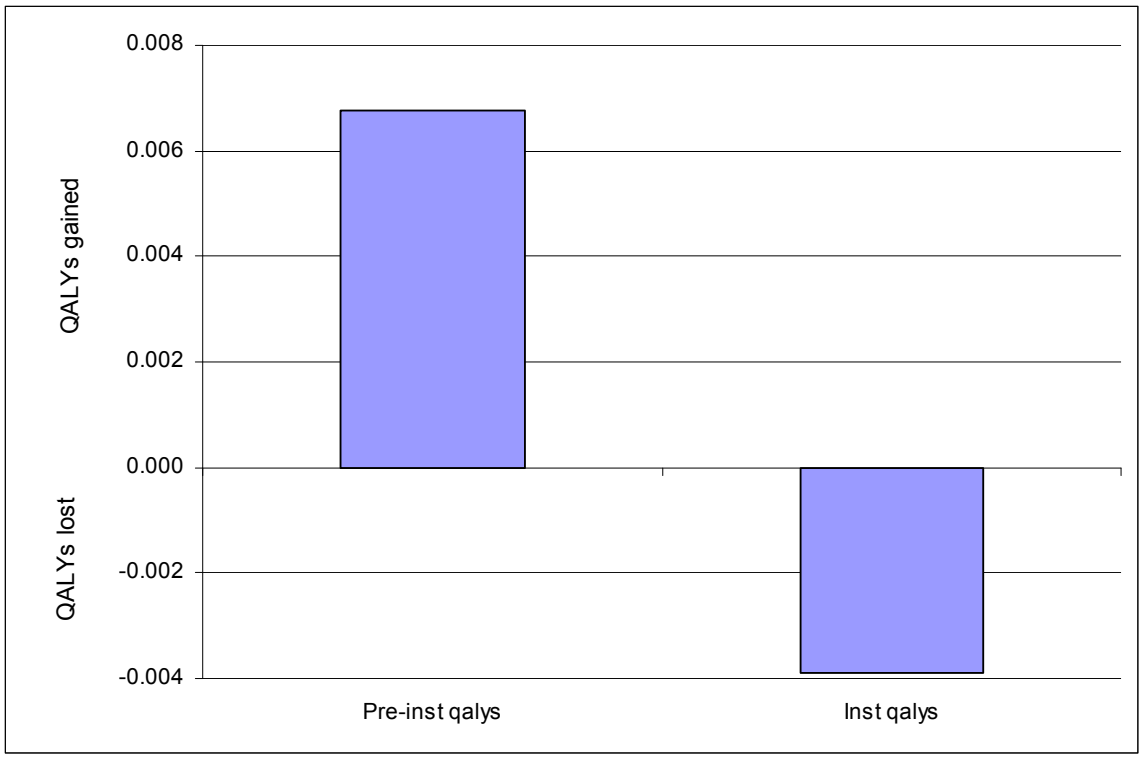


FIGURE 90 *Base-case QALY components of memantine compared to best supportive care for moderate to severe Alzheimer's disease*



7.4.2.3. Summary of probabilistic sensitivity and deterministic analysis

As with the AChEIs, there is a great deal of parameter uncertainty in the cost–utility of memantine compared to best supportive care. However, there is a low probability of memantine being cost-effective treatment for any willingness to pay threshold.

7.4.2.4. One-way sensitivity analysis

Treatment effect on mortality

Assuming a treatment effect on survival leads to a mean estimate of overall survival of 42.1 months for best supportive care and 42.7 months for treatment with memantine: an additional 18 days of life for individuals receiving memantine compared to best supportive care. It is estimated that treatment with memantine provides an additional 0.016 QALYS compared to best supportive care over a patient's lifetime when a treatment effect on survival is assumed.

However, these QALY gains cost an additional £1,738 leading to a cost per QALY of £107,900 for memantine compared to best supportive care (see *Table 123*).

TABLE 123 *Incremental cost–utility analysis for moderate to severe Alzheimer's disease when a treatment effect on survival is assumed*

Treatment	Costs	QALYs	Incremental costs	Incremental QALYs	ICER[a]
Best supportive care	£78,136	1.214			
Memantine (20mg)	£79,874	1.231	£1,738	0.016	£107,900

a Cost per QALY rounded to the nearest £100

The assumption of a treatment effect on survival leads to a lower cost per QALY than the base-case analysis. Examination of the cost components in *Figure 91* reveals that there are no cost savings associated with memantine over best supportive care. However, *Figure 92* demonstrates that there are QALY gains in both states, pre-institutionalized and institutionalized. This is in contrast to QALY losses in the institutionalized state in the base-case analysis (refer back to *Figure 90*), since longer time is spent in the institutionalized state when a treatment effect on survival is assumed. The reduced cost per QALY compared to the base-case analysis is due to the extra QALY gains from being in the institutionalized state which are not countered by the additional costs of a longer time spent in the institutionalized state.

FIGURE 91 *Cost components for memantine compared to best supportive care assuming a treatment effect on survival*

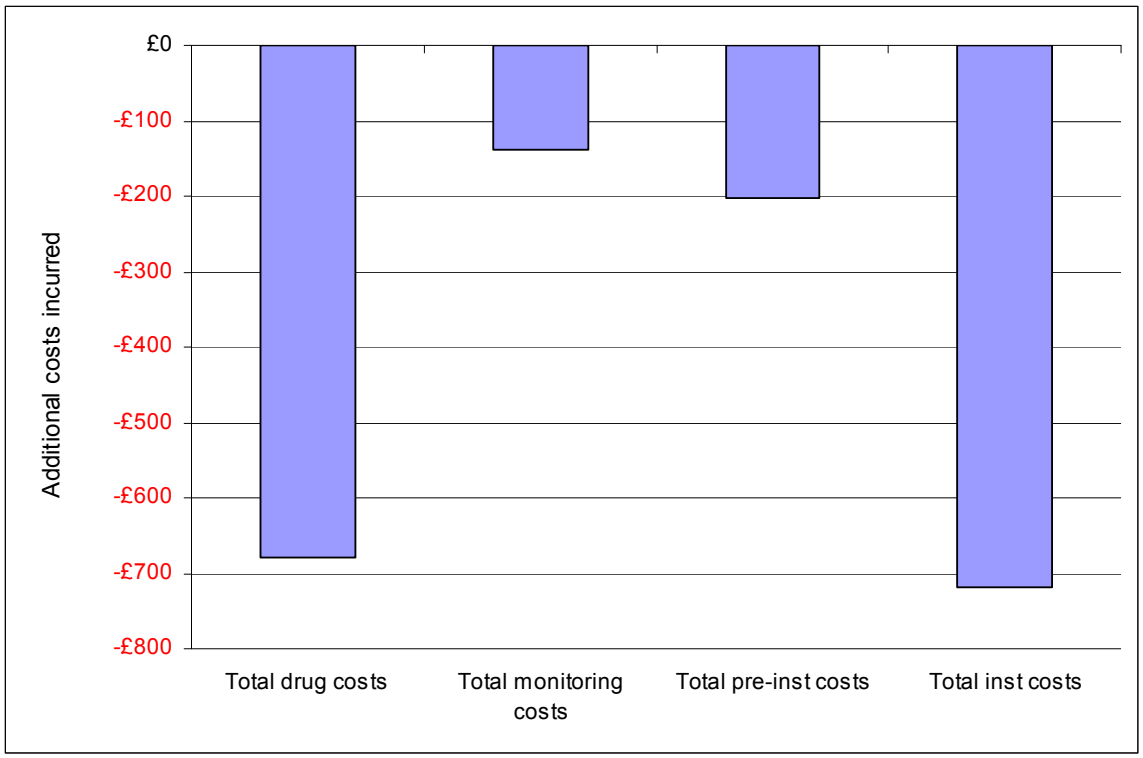
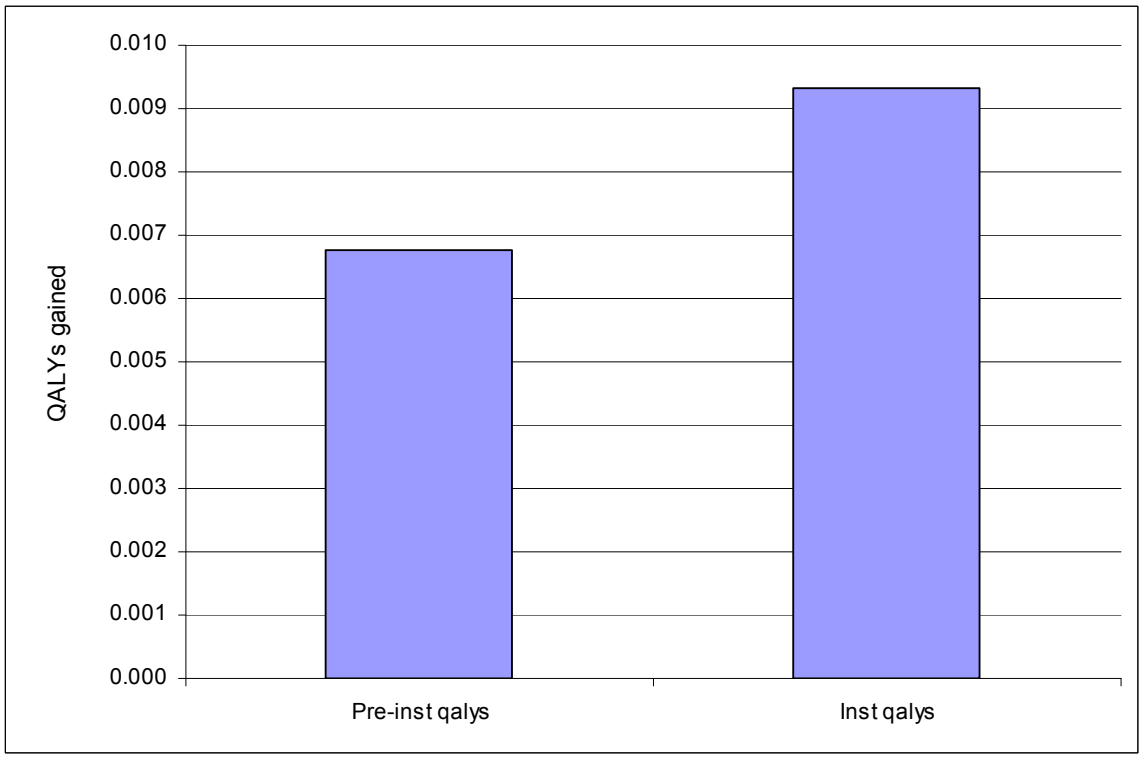


FIGURE 92 *QALY components for memantine compared to best supportive care assuming a treatment effect on survival*



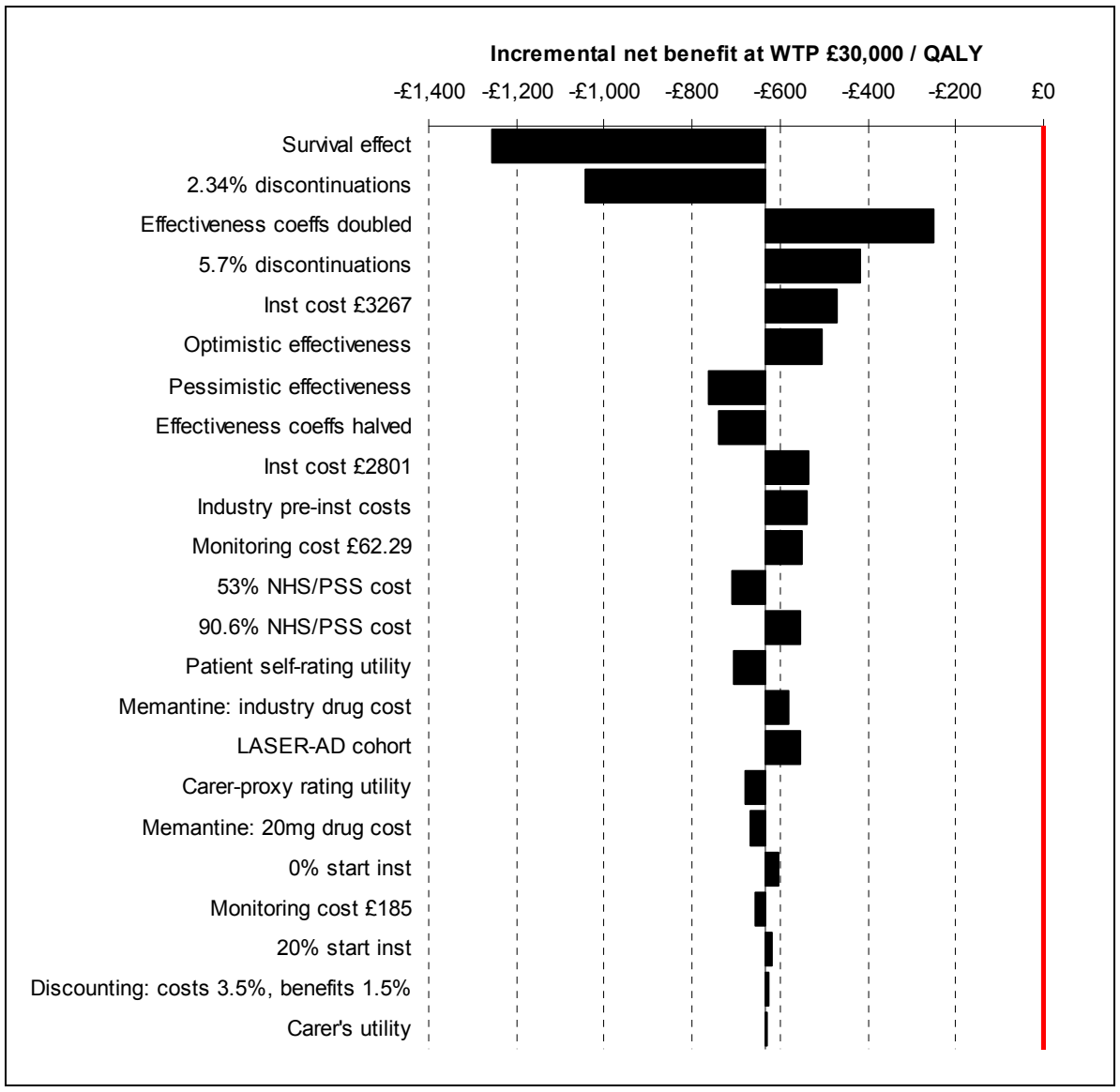
Further one-way sensitivity analyses

As with the cholinesterase inhibitors, a number of one-way sensitivity analyses have been undertaken to assess important assumptions and parameters in the model. The same assumptions as those outlined in *Table 120* above are applied to the memantine dataset, with additional sensitivity analyses outlined in *Table 124*. A tornado plot showing the impact on the cost-effectiveness of changing individual parameters and assumptions is given in *Figure 93.*

TABLE 124 *Additional parameter and assumption changes for deterministic sensitivity analyses for base-case analysis of memantine with moderate to severe Alzheimer's disease*

Parameter/ assumption	Base-case	Deterministic sensitivity analysis	Reference in tornado plots
Drug costs	See *Table 113*	Industry cost for memantine; 20mg cost for memantine	Drug cost
Severity of cohort	Mild to moderate	Moderate; Severe	Cohort severity
% start in nstitutional care	40%	20%	% start inst

FIGURE 93 *One-way sensitivity analyses for the incremental net benefit of memantine compared to best supportive care*



Note that none of the one-way sensitivity analyses led to a positive net benefit of memantine compared to best supportive care. The pattern of the importance of the assumptions is very similar to the tornado plots for the AChEIs. Readers are therefore referred back to *Table 121* which explains the importance of the assumptions. The assumption of a survival effect has the largest impact on the net benefit from the base-case analysis of memantine in moderate to severe Alzheimer's disease, along with the assumptions of lower or higher discontinuations.

7.4.2.5. Summary of one-way sensitivity analysis

There are many uncertainties in the PenTAG model for treatment with memantine in people with Alzheimer's disease. The assumption of a survival effect with treatment has the largest impact on the cost-effectiveness findings. As noted above, there is no direct evidence from RCTs that memantine extends survival, however memantine does influence the covariates explaining some of the variation in overall survival.

7.4.3. Exploratory subgroup cost–utility analyses

Exploratory subgroup analyses were undertaken to assess

- Decision problem 1b in *Table 103*: treatment of mild Alzheimer's disease with AChEIs
- Decision problem 1c in *Table 103*: treatment of moderate Alzheimer's disease with AChEIs
- Decision problem 2a in *Table 103*: treatment of moderate Alzheimer's disease with memantine
- Decision problem 3 in *Table 103*: treatment of moderate Alzheimer's disease with AChEIs or memantine

As noted in Section *7.3.3.1*, caution should be used in the interpretation of these results as the effectiveness estimates used are not restricted to the severities assessed. That is, they have not been derived from trials which have recruited patients of that disease severity or from trial sub-group analyses. Therefore, the main differences between these analyses and the base-case analyses are the baseline population characteristics. Refer back to *Table 105* for the age, MMSE and ADL parameter values for the different AD severities used in the following exploratory analyses.

7.4.3.1. Treatment of mild Alzheimer's disease (Decision problem 1b)

The results of an explorative cost–utility analysis of AChEIs for a cohort of people starting the model with mild Alzheimer's disease are presented in *Table 125*. Rivastigmine patches are

estimated to have the lowest ICER (£81,700 per QALY), with rivastigmine capsules and donepezil dominated.

TABLE 125 *Cost–utility results of AChEI use in people with mild Alzheimer's disease*

Treatment	Costs	QALYs	Incremental costs	Incremental QALYs	ICER[ab]
Best supportive care	£75,515	1.749			
Rivastigmine patch (10cm2)	£76,068	1.755	£553	0.007	£81,700
Galantamine (16-24mg)	£76,092	1.756	£24	0.0001	£178,000
Donepezil (10mg)	£76,210	1.756	Dominated		
Rivastigmine capsules (9-12mg)	£76,261	1.755	Dominated		

a Rounded to nearest £100

b Compared to next cheapest, non-dominated technology

7.4.3.2. Treatment of moderate Alzheimer's disease (Decision problems 1c and 3)

The results of an explorative cost–utility analysis of AChEIs or memantine for a cohort of people starting the model with moderate Alzheimer's disease are presented in *Table 126.* Memantine is dominated and so the results presented in *Table 126* address both decision problem 1c and 3. Rivastigmine patches are estimated to have the lowest ICER (£58,000 per QALY), with rivastigmine capsules and donepezil dominated. Note that the total costs associated with treatment of a moderate cohort are smaller than those for a mild cohort, due to less time spent on treatment. The total QALYs are also smaller for the moderate than the mild group as survival is lower and disease severity is greater.

TABLE 126 *Cost–utility results of treatment in people with moderate Alzheimer's disease*

Treatment	Costs	QALYs	Incremental costs	Incremental QALYs	ICER[ab]
Best supportive care	£67,536	1.500			
Rivastigmine patch (10cm2)	£67,999	1.508	£463	0.008	£58,000
Galantamine (16-24mg)	£68,021	1.508	£22	0.0002	£147,900
Donepezil (10mg)	£68,145	1.508	Dominated		
Rivastigmine capsules (9-12mg)	£68,198	1.507	Dominated		
Memantine (15-20mg)	£68,069	1.505	Dominated		

a Rounded to nearest £100

b Compared to next cheapest, non-dominated technology

7.4.3.3. Treatment of severe Alzheimer's disease (decision problem 2b)

The results of an explorative cost–utility analysis of memantine for a cohort of people starting the model with severe Alzheimer's disease are presented in *Table 127*. The resultant ICER of £279,700 per QALY is slightly higher than that for the cohort of people with moderate to severe Alzheimer's disease, as there are slightly fewer QALYs gained when assuming a severe cohort than a moderate cohort. However, this difference is so small it is not clear from comparison of *Table 122* with *Table 127* due to the decision to restrict reporting of QALYs to three decimal places. As pointed out, the data informing the effectiveness of memantine in this severe cohort is from a trial where the participant population ranged from moderate to severe Alzheimer's disease. Therefore, these results should be treated with caution, as with the results presented above in *Table 125* and *Table 126.*

TABLE 127 *Cost–utility results of memantine in people with severe Alzheimer's disease*

Treatment	Costs	QALYs	Incremental costs	Incremental QALYs	ICER[a]
Best supportive care	£67,993	1.012			
Memantine (15-20mg)	£68,694	1.014	£701	0.003	£279,700

a Rounded to the nearest £100

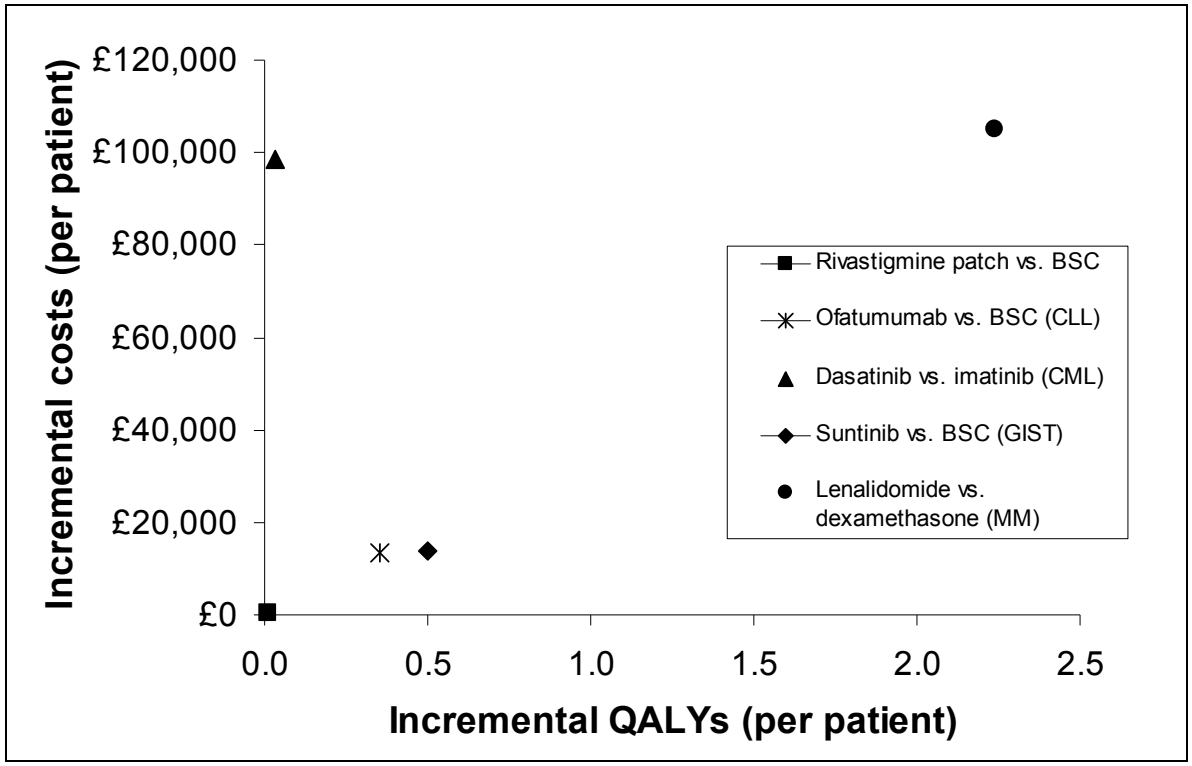
7.5. Summary of cost-effectiveness findings

The cost–utility results for AChEIs in people with mild to moderate Alzheimer's disease (Section *7.4.1*) and memantine in people with moderate to severe Alzheimer's disease (Section *7.4.2*) indicate a great deal of uncertainty, only some of which is expressed in the PSA. When considering the AChEIs, there is a 43% probability that best supportive care is *not* the most cost-effective treatment option at a willingness to pay of £30,000 per QALY (and 38% at a willingness to pay of £20,000 per QALY).

It is important to note that the QALY gains and additional costs for a particular AChEI over best supportive care or another AChEI are both very small. To demonstrate these low incremental costs and QALYs, *Figure 94* (below) allows a comparison of the base-case incremental costs and QALYs from the current assessment of Alzheimer's drugs (results for rivastigmine patches) with the base-case incremental costs and QALYs estimated from recent MTAs conducted by the technology assessment group and recent STAs from industry. Whereas previous assessments have estimated typical incremental QALY gains of 0.5 or

more at an incremental cost of £10,000 or more, the current model-based assessment estimates incremental QALY gains of <0.01 with incremental costs <£800.

FIGURE 94 *Cost-effectiveness plane for the base-case estimate for rivastigmine patches in the current cost–utility analysis and base-case estimates from recent MTAs and STAs*



Source: www.nice.org.uk [cited 2010 June 11]

The probability that memantine is cost-effective in a moderate to severe cohort compared to best supportive care (Section *7.4.2*) at a willingness to pay of £30,000 per QALY is <4% (and 2.6% at a WTP of £20,000 per QALY). This probability increases little as the willingness to pay threshold increases, and under no scenarios did the net benefit for memantine compared to best supportive care become positive at a willingness to pay of £30,000.

Base-case deterministic and probabilistic ICERs for treating mild to moderate and moderate to severe Alzheimer's disease are presented in

TABLE 128.

TABLE 128 *Base-case deterministic and probabilistic ICERs for treatment of mild to moderate and moderate to severe Alzheimer's disease*

	Mild to moderate AD		Moderate to severe AD	
	Deterministic	**Probabilistic**	**Deterministic**	**Probabilistic**
Rivastigmine patches (10cm^2)	£61,100[a]	£59,800[a]	NA	NA
Galantamine (16–24mg)	£151,100[b]	£157,800[b]	NA	NA
Donepezil (10mg)	Dominated	Dominated	NA	NA
Rivastigmine capsules (9-12mg)	Dominated	Dominated	NA	NA
Memantine (15-20mg)	NA	NA	£248,500	£288,800

[a] Compared to best supportive care, the next cheapest non-dominated technology

[b] Compared to rivastigmine patches, the next cheapest non-dominated technology

7.6. Comparison of PenTAG model with SHTAC model

Table 129 shows there is very little difference in the cost–utility estimates of donepezil and galantamine provided by the PenTAG model and those presented by SHTAC in the HTA related to the previous assessment. The difference in the cost–utility of rivastigmine capsules is likely due to differences in monthly costs of the drug which was £74 in the 2004 report and estimated to be £98 in the current review.

TABLE 129 *ICERs[a] from the PenTAG model and the SHTAC model for AChEIs compared to best supportive care in people with mild to moderate Alzheimer's disease*

	PenTAG model	SHTAC model		
	Deterministic base-case results	**As reported in HTA monograph[145]**	**3.5% discount rates; probabilities as probabilities; 20 yr time horizon**	**As previous column plus discontinuations and PenTAG effectiveness and cost estimates**
Donepezil (10mg)	£80,400	£80,900	£66,500	£45,300
Galantamine (16 – 24mg)	£62,700	£68,000	£55,000	£37,700
Rivastigmine capsules (9-12mg)	£100,600	£58,000	£46,100	£72,200
Rivastigmine patches ($10cm^2$)[b]	£61,100			

[a] Rounded to nearest £100

[b] Only rivastigmine capsules were evaluated in the SHTAC model, not the patches

We have modified the estimates of the SHTAC model to account for current discount rates (3.5% for both costs and benefits), a longer time horizon (20 years) and have amended the SHTAC model so that all probabilities are represented as probabilities (and not hazards as done in the previous review). These updated estimates are also shown in *Table 129* (third column), alongside further updated estimates from the SHTAC model for the inclusion of treatment discontinuations and updated effectiveness and cost parameters (see *Table 130* for comparison of effectiveness and cost parameter inputs between SHTAC and PenTAG models). The results from the SHTAC model using current cost and effectiveness estimates (final column) are lower than those from the PenTAG model and this is likely to be due to a number of factors, which are now discussed in light of the outputs from the two models (see *Table 131* for comparison of outputs for the cost-effectiveness of donepezil from the PenTAG model and the updated SHTAC model). However, note that the incremental costs and QALYs are small for both the SHTAC and PenTAG models, which implies that the ICERs are volatile.

TABLE 130 *Effectiveness and cost inputs from SHTAC and PenTAG models*

Parameter	SHTAC 2004 value	PenTAG 2010 value
ADAS-cog effectiveness		
Donepezil	3.01	2.90
Rivastigmine	3.08	2.34
Galantamine	3.28	3.05
Monthly drug costs		
Donepezil	£97	£97
Rivastigmine	£74	£98
Galantamine	£91	£83
6-monthly monitoring visit cost	£108	£158
Monthly pre-FTC/inst cost	£328	£2051
Monthly FTC/inst cost	£937	£2117 (£2941 * 72%)

TABLE 131 *Comparison of outputs from PenTAG model and updated SHTAC model for donepezil for mild to moderate cohort*[a]

		Model outputs		Incremental values	
Output	**Treatment**	**Updated SHTAC**[b]	**PenTAG**	**Updated SHTAC**[b]	**PenTAG**
ICER		£45,297	£80,377		
Total costs	Donepezil	£142,742	£70,863		
	No treatment	£140,861	£70,237	£1,881	£627
Total QALYs	Donepezil	2.548	1.592		
	No treatment	2.506	1.584	0.042	0.008
Undiscounted total life years		5.616	3.836		
Undiscounted life years in community	Donepezil	2.581	2.431		
	No treatment	2.403	2.401	0.178	0.03
Undiscounted years in institutional care	Donepezil	3.916	1.405		
	No treatment	4.094	1.436	-0.178	-0.03
Mean treatment duration (years)		0.87[a]	0.73		
Total drug costs		£1,587	£779		
Total monitoring costs	Donepezil	£431	£212		
	No treatment	£0	£0	£431	£212
Total pre-inst costs	Donepezil	£63,460	£40,237		
	No treatment	£59,145	£39,954	£4,315	£283
Total inst costs	Donepezil	£99,510	£29,635		
	No treatment	£104,041	£30,282	-£4,531	-£647

a All costs and QALYs discounted

b Updated SHTAC model: discount rates of 3.5% for both costs and benefits, a time horizon of 20 years, inclusion of treatment discontinuations and updated effectiveness (i.e. ADAS-cog estimate) and cost parameters

There is a very large difference between the total costs in the SHTAC model compared to the PenTAG model, with the SHTAC model having much larger institutional costs. This can be attributed to the SHTAC model estimating greater survival than the PenTAG model (5.6 years from the SHTAC model compared to 3.8 years in the PenTAG model). However, there is little difference in the time to FTC/institutionalization (for the donepezil individuals this is 2.58 in the SHTAC model and 2.43 in the PenTAG model). Therefore, the SHTAC model estimates a longer time spent in FTC/institutionalization than the PenTAG model.

Furthermore, a larger difference in incremental costs between the donepezil and best supportive care groups is seen with the SHTAC model compared to the PenTAG model, especially for total institutional care costs: £4,531 from the SHTAC model compared to £647 in the PenTAG model. This is attributed to the larger delay in FTC/institutionalization for treated individuals estimated by the SHTAC model compared to the PenTAG model. The

SHTAC model predicts a delay of about 2 months (0.178 years) for donepezil treated individuals while the PenTAG model only estimates a delay to institutionalization of 11 days (0.03 years). This is also reflected in the QALYs with the SHTAC model estimating an incremental 0.042 QALYs between donepezil and best supportive care and the PenTAG model only estimating incremental QALYs of 0.008. This leads to the SHTAC model predicting an ICER of £45,300 per QALY compared to an ICER of £80,800 per QALY from the PenTAG model.

Although there are still many differences in the overall structure between the PenTAG model and the updated SHTAC model, e.g. there is just a single cost and utility assigned in the pre-FTC state in the SHTAC model, while in the PenTAG model, costs and utilities pre-institutionalization are dependent on severity, the main difference between the models is the delay to FTC/institutionalization for treated individuals. This is estimated to be much larger from the SHTAC model compared to the PenTAG model.

The data used to predict FTC/institutionalization differs between the two models. The SHTAC model used a US study where ADAS-cog, psychiatric symptoms, extrapyramidal symptoms, age at onset and duration of illness were identified to be statistically significant predictors for time to full-time care. The PenTAG model used UK data where age was found to be a significant predictor for time to institutionalization, but to incorporate a treatment effect, MMSE and Barthel-ADL were included even though they were not found to be significant predictors of time to institutionalization (see *Figure 64*a and b, and *Figure 65*a and b). The estimated delay in institutionalisation is greater in the SHTAC model than in the PenTAG model since the treatment effect estimates used in the SHTAC model were relatively larger and the effectiveness estimates have a larger impact on disease progression. A comparison of the study by Wolstenholme and colleagues and the study by Stern and colleagues used to predict time to institutionalization/full-time care in the PenTAG and SHTAC models respectively is available in *Table 132*. Although the Wolstenholme study is UK-based and has longer follow-up than the US study, it is only based on 92 individuals. Thus, there are concerns that this dataset may not fully represent the target population of England and Wales.

TABLE 132 *Comparison of Alzheimer's disease progression datasets: Stern and colleagues (1997)[191], Wolstenholme and colleagues (2002)[181] and Livingston and colleagues[182]*

Study characteristic	Stern et al[191] dataset	Wolstenholme et al[181] dataset	LASER-AD study[182]
Geographical setting	US: New York, Baltimore and Boston	UK: Oxfordshire	UK: North London and Essex
Event definition	Requiring full-time care; "equivalent institutional care"	Institutionalization	Entering 24-hour care
AD sample size	236	92	224
Available care cost estimates	No	Yes	Yes
Start data collection	Not reported	1988/9	Not reported
Length of follow-up	Up to 7 years	Up to 11 years	Up to 4.5 years
Predictors of time to event (and stat significance)	Modified MMSE (p<0.1); Psychosis (p<0.1) ; Age at onset (p<0.1); Extrapyramidal symptoms (p<0.1); Duration of illness (p<0.1)	MMSE (ns) Barthel-ADL (ns) Age (p=0.0009)	MMSE (p=0.001); Hours spend caring (p=-0.03); Level of education (p=0.004); Relationship to carer (partner v. family p=0.04; partner v. paid, p= 0.001; family v. paid, p=0.001)
Mean age at study entry	73.1 (SD 8.9) years	78.1 (SD 6.9) years	81 (SD 7.4) years
Average severity at study entry	Mild at study entry (MMSE>15)	MMSE = 14.4 (SD 6.7)	30% MMSE<15 40% 14<MMSE<20 30% MMSE>19
Time since onset/diagnosis	Average time since onset (NOT diagnosis): 3.9 years	Average time since diagnosis: 4.9 years	Unclear

7.7. Comparison of PenTAG model with industry models

7.7.1. Eisai/Pfizer v. PenTAG: donepezil

The base-case results of the Eisai/Pfizer model for both a moderate cohort and a mild cohort suggest that donepezil dominates when compared to no treatment. However, the base-case analyses from the PenTAG model for a moderate cohort and a mild cohort indicate that donepezil has an ICER of £77,400 and £102,000 / QALY, respectively, when compared to best supportive care. To allow comparison between the PenTAG and Eisai/Pfizer models it is assumed that the Eisai/Pfizer definition of no treatment and the PenTAG definition of best supportive care are equivalent.

The Eisai/Pfizer model structure and simulation method is very different to the structure of the PenTAG model, and this hinders systematic comparison between the two models. However, to allow initial comparison between the base-case results of the two models, outputs from these models for a moderate cohort are presented in *Table 133*, and those for a

mild cohort are presented in *Table 134*. Note that in the Eisai/Pfizer base-case analysis, carer's utility is included alongside patient utility. Only patient utility is included in the PenTAG base-case model. Also note that the Eisai/Pfizer model predicts a *shorter* survival for the mild cohort than for the moderate cohort. This inconsistency was noted by the DSU in Section 6.3.

TABLE 133 *Outputs from PenTAG and Eisai/Pfizer models for donepezil (moderate cohort)*[a]

		Model outputs		Incremental values	
Output	**Treatment**	**Eisai/Pfizer**	**PenTAG**	**Eisai/Pfizer**	**PenTAG**
ICER		Donepezil dominates	£77,428		
Total costs	Donepezil	£102,086	£68,145		
	No treatment	£103,969	£67,536	-£1,883	£609
Total QALYs	Donepezil	4.353 (patient + carer)[b]	1.508 (patient only)		
	No treatment	4.245 (patient + carer)[b]	1.500 (patient only)	0.108	0.008
Undiscounted total life years		4.603	3.633		
Undiscounted life years in community	Donepezil	1.852	2.307		
	No treatment	1.685	2.276	0.167	0.031
Undiscounted years in institutional care	Donepezil	2.751	1.326		
	No treatment	2.918	1.357	-0.167	-0.031
Mean treatment duration (years)		1.89	0.73		
Total drug costs		£1,973	£768		
Total monitoring costs	Donepezil	£208	£209		
	No treatment	£0	£0	£208	£209
Total pre-inst costs	Donepezil	£39,201	£39,003		
	No treatment	£37,413	£38,709	£1,788	£294
Total inst costs	Donepezil	£60,705	£28,165		
	No treatment	£66,556	£28,827	-£5,851	-£662

a All costs and QALYs discounted

b Eisai/Pfizer base-case includes carer QALYs, therefore total QALYs = patient QALYs + carer QALYs. Donepezil total QALYs = 1.332 + 3.021; No treatment total QALYS = 1.234 + 3.011

TABLE 134 *Outputs from PenTAG and Eisai/Pfizer models for donepezil (mild cohort)*[a]

		Model outputs		Incremental values	
Output	**Treatment**	**Eisai/Pfizer**	**PenTAG**	**Eisai/Pfizer**	**PenTAG**
ICER		Donepezil dominates	£101,703		
Total costs	Donepezil	£79,023	£76,210		
	No treatment	£82,409	£75,515	-£3,386	£695
Total QALYs	Donepezil	4.267 (patient + carer)[b]	1.756 (patient only)		
	No treatment	4.120 (patient + carer)[b]	1.749 (patient only)	0.147	0.007
Undiscounted total life years		4.110	4.243		
Undiscounted life years in community	Donepezil	2.161	2.669		
	No treatment	1.926	2.642	0.235	0.027
Undiscounted years in institutional care	Donepezil	1.949	1.574		
	No treatment	2.184	1.600	-0.235	-0.027
Mean treatment duration (years)		2.23	0.76		
Total drug costs		£2,281	£798		
Total monitoring costs	Donepezil	£240	£217		
	No treatment	£0	£0	£240	£217
Total pre-inst costs	Donepezil	£37,938	£42,441		
	No treatment	£37,128	£42,205	£810	£236
Total inst costs	Donepezil	£38,564	£32,754		
	No treatment	£45,282	£33,310	-£6,718	-£556

a All costs and QALYs discounted

b Eisai/Pfizer base-case includes carer QALYs, therefore total QALYs = patient QALYs + carer QALYs. Donepezil total QALYs = 1.502 + 2.765; No treatment total QALYS = 1.370 + 2.750

The general pattern in the Eisai/Pfizer outputs is similar for the moderate and mild cohorts, as it is for the PenTAG model. Therefore, the following commentary only relates to comparison of the Eisai/Pfizer and PenTAG models for the moderate cohort, although the same general trends in the differences between the models are seen for the mild cohort. For a moderate cohort, there is longer overall survival in the Eisai/Pfizer model (undiscounted life years of 4.60 v. 3.63). However, the Eisai/Pfizer model predicts a shorter time living in the community than the PenTAG model (1.85 v. 2.31 years for donepezil treated patients). In

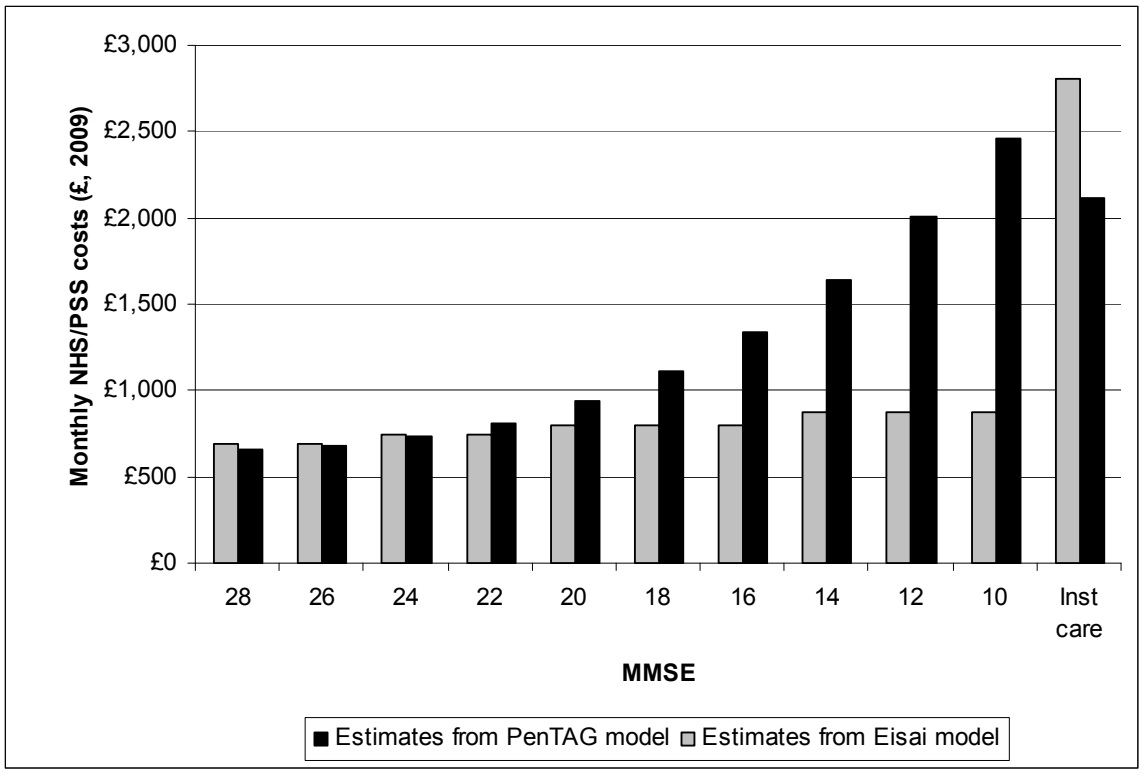
the Eisai/Pfizer model 40% of the remaining lifetime of patients is estimated to be in the community, compared with 64% of the remaining lifetime of patients in the PenTAG model.

Conversely therefore, the Eisai/Pfizer model estimates a longer time spent in institutional care than estimated by the PenTAG model. This is reflected in the greater total costs estimated for being in institutional care. The costs associated with being in institutional care are the origin of the main cost differences between donepezil and no treatment, with a difference of £5,851 from the Eisai/Pfizer model and £662 from the PenTAG model (see last row of *Table 133*). These results show a greater delay to institutional care with treatment estimated by the Eisai/Pfizer model: a difference of about two months (0.17 years) from the Eisai/Pfizer model compared to a difference of around 11 days (0.03 years) from the PenTAG model. This difference in treatment effect is multiplied by a greater assumed increase in care costs when moving from the community into institutionalised care (over £1,800 per month; see below)

The greater assumed treatment effect in the Eisai/Pfizer model results in a greater incremental QALY gain in the Eisai/Pfizer model compared to the PenTAG model. In both models, the delay to institutionalization leads to increased overall QALYs as utilities in pre-institutionalization are greater than utilities in the institutionalized state. Thus, given that donepezil is then estimated to be cost saving and generates more QALYs than best supportive care in the Eisai/Pfizer model, it dominates best supportive care.

The other differences between the two models are the cost inputs, particularly the NHS/PSS care costs in the community and in institutional care. The Eisai/Pfizer pre-institutionalized care costs are reported by MMSE, while the PenTAG pre-institutionalized care costs are calculated by time to institutionalization. Using the equation described in Section *7.3.9.1* to relate MMSE to time to institutionalization, it is possible to compare the community-living costs from each model defined by MMSE. As can be seen clearly from *Figure 95,* the community costs assumed in the PenTAG model are much larger than the assumed Eisai/Pfizer costs and allow for more change in costs as individuals progress over time. However, there is some concern that the community costs used by Eisai/Pfizer have not been appropriately translated from the CDR scale to the MMSE scale.

FIGURE 95 *Monthly NHS/PSS costs by MMSE for individuals living in the community from the Eisai/Pfizer model and the PenTAG model*



In their industry submission, Eisai/Pfizer cite Knapp and colleagues[6] who report the annual cost of community care by CDR. To obtain their cost estimates Eisai/Pfizer have assumed that mild on the CDR scale is equivalent to MMSE>25, moderate on the CDR scale is equivalent to MMSE > 15 and MMSE <20, and severe on the CDR scale is equivalent to MMSE < 10. To calculate community costs of care for the remaining severities defined by Eisai/Pfizer, 19 < MMSE <26 and < 9 MMSE <16, they have interpolated the mean cost values from the adjacent severities. Perneczky and colleagues[214] indicate that the severities defined by CDR mild, moderate and severe are a good approximation to the MMSE severities of mild (MMSE 25-21), moderate (MMSE 20-11) and severe (MMSE 10-0). These approximations do not relate as expected to the Eisai/Pfizer cost estimates and there is no indication of any other published evidence used by Eisai/Pfizer to approximate CDR and MMSE scores.

The monthly costs of institutional care with the Eisai/Pfizer model assume costs of £2801 per month while the PenTAG model assumes costs of £2117 per month (accounting for the 28% of institutional costs assumed to be privately funded). Again, there is concern that the

monthly costs for institutional care in the Eisai/Pfizer model are not solely costs funded by the NHS/PSS.

In both models a six-monthly out-patient monitoring visit is accounted for. Eisai/Pfizer describe this as a geriatrician visit, while it is described as a consultant led out-patient visit for the PenTAG model. The cost for this monitoring visit is very different between the Eisai/Pfizer and PenTAG models: £62 from Eisai/Pfizer v. £158 from PenTAG. Both costs are cited from the National Schedule of Reference Costs (service code 430 Geriatric Medicine), with the Eisai/Pfizer cost cited as 2007-8 and the PenTAG cost as 2008-9. There are also slight differences in the daily drug costs for 10mg donepezil between the two models. The Eisai/Pfizer submission reports a daily cost of £3 from the NHS drug tariff, compared to £3.18 used in the PenTAG model from the BNF58.

Inputting the above Eisai/Pfizer cost estimates (for the NHS/PSS care costs (community and institution), the 6-monthly monitoring visit and the daily drug costs) into the PenTAG model gives an ICER of £28,600 compared to the PenTAG base-case estimate of £77,400 for the moderate cohort. This is mainly the cumulative effect of the different community and institutional care costs and costs of the monitoring visit between the Eisai/Pfizer and PenTAG model. Investigation into the basis of Eisai/Pfizer's cost of £62 for the geriatrician visit could not identify the exact source of this value. In fact, no appropriate cost values below £100 for the monitoring visit were identified. The cost of £2801 for institutional care in the Eisai/Pfizer model was found to include costs not funded by the NHS/PSS. Applying the assumption that 28% of institutional care costs are privately funded, a re-analysis of the PenTAG model with Eisai/Pfizer cost estimates assuming only 72% of institutional care costs are funded by the NHS/PSS gives an ICER of £59,800. This value is quite different to the ICER of £28,600 when all institutional care costs are assumed to be funded by the NHS/PSS.

Even with alternative cost estimates, the PenTAG model does not lead to an estimate of donepezil dominating best supportive care. As pointed out above, the largest difference between the models comes from the treatment effect. It is difficult to fully assess and evaluate the differences in the effectiveness evidence used by Eisai/Pfizer and by PenTAG because of differences in the model structure, the software used to run the Eisai/Pfizer model and the fact that information regarding the derivation of the effectiveness estimates was not fully reported in the Eisai/Pfizer submission.

7.7.2. Lundbeck v. PenTAG: memantine

The structure of the Lundbeck and PenTAG base-case models are similar, however there are a number of important differences which are shown in *Table 135*. The assumptions made in the Lundbeck submission of moderate severity at the start of the model and that treatment stops once patients enter FTC are the same as in the PenTAG model of memantine for a cohort with moderate AD at the start of the model. This differs to the PenTAG base-case for memantine which assumes a moderate to severe cohort start the model and that treatment continues even in institutional care. Therefore, the Lundbeck model is compared to the PenTAG memantine model for a cohort with moderate disease severity at the start.

TABLE 135 *Differences in model structure and parameter values between the Lundbeck and PenTAG models for moderate to severe Alzheimer's disease*

	Lundbeck model	**PenTAG model**
Modelling assumption		
Definition of alive states	Pre-FTC/FTC[a]	Pre-institutionalization/ institutionalization[b]
Data describing AD progress	LASER-AD	IPD from Wolstenholme et al
Severity at start of model	Moderate	Moderate to severe
Treatment stopping rules	Stop when in FTC	Continue until death or treatment discontinuations
Parameter value		
Monthly pre-FTC/institutinalized care costs	£724	Dependent on severity, >£724.
Monthly FTC/institutional care costs	£3201	£2117
Monthly memantine drug costs	£64.80	£71.28

a FTC defined as a patient becoming either dependent or institutionalized

b Institutionalization defined as living in residential or hursing care, or in a hospital on a long-term or permanent basis

The results from the two models differ markedly, with Lundbeck's model estimating that memantine dominates while the PenTAG model estimates an ICER of £103,900 for a moderate cohort. Results from the Lundbeck and PenTAG models for treatment with memantine are shown in *Table 136*.

TABLE 136 *Comparison of outputs from PenTAG and Lundbeck models for memantine compared to best supportive care*[a]

		Model outputs		Incremental values	
Output	**Treatment**	**Lundbeck**	**PenTAG**	**Lundbeck**	**PenTAG**
ICER		Memantine dominates	£103,885		
Total costs	Memantine	£93,076	£68,069		
	No treatment	£94,787	£67,536	-£1,711	£533
Total QALYs	Memantine	1.533	1.505		
	No treatment	1.502	1.500	0.031	0.005
Total pre-inst/FTC QALYs	Memantine	0.870	1.136		
	No treatment	0.813	1.125	0.057	0.011
Total inst/FTC QALYs	Memantine	0.661	0.369		
	No treatment	0.690	0.374	-0.029	-0.005
Expected overall survival (years)		3.7	3.633		
Expected time to FTC/ institutional care (years)	Memantine	1.73	2.296		
	No treatment	1.65	2.276	0.08	0.02
Time in FTC/institutional care	Memantine	1.97	1.336		
	No treatment	2.05	1.357	-0.08	-0.02
Mean treatment duration		1.73	0.66		
Total drug costs		£1,348	£565		
Total monitoring costs	Memantine	£106	£209		
	No treatment	£0	£0	£106	£209
Total pre-inst costs	Memantine	£16,642	£38,901		
	No treatment	£14,324	£38,709	£2,318	£192
Total inst costs	Memantine	£77,133	£28,394		
	No treatment	£80,464	£28,827	-£3,331	-£433

a All costs and QALYs discounted

Overall survival is estimated to be similar between the models (3.7 years from the Lundbeck model and 3.6 years from the PenTAG model), but time spent in pre-FTC/institutionalization is greater in the PenTAG model (for memantine 1.73 years in pre-FTC from the Lundbeck model compared to 2.30 years in the PenTAG model). As with the comparison of the PenTAG and Eisai/Pfizer models, memantine is assumed to have a greater treatment effect (expressed as time to FTC/institutionalization) in the Lundbeck model compared to the PenTAG model. In the Lundbeck model memantine is estimated to delay FTC by about a month (0.08 years), while in the PenTAG model the estimated delay to institutionalization in a

moderate cohort is only about 7 days. Moreover, substantially greater costs are attributed to the FTC state in the Lundbeck model compared to the institutionalized state in the PenTAG model (£3267 per month in Lundbeck compared to £2117 per month in the PenTAG model). The Lundbeck model assumes a single cost for pre-FTC of £724. This is of similar magnitude to the PenTAG pre-institutionalized care costs for those with MMSE > 21 in the pre-institutionalized state, but much lower than the costs for individuals assumed to have lower MMSE scores in the PenTAG model (see *Figure 95*). This means that delaying FTC/institutionalization has a greater impact on costs in the Lundbeck model than in the PenTAG model.

The PenTAG model was re-run assuming the institutional care costs of £3,267, and a single pre-institutionalized care cost of £724. These changes produce an estimated ICER of £49,563 from the PenTAG model. However, assuming that only 72% of FTC/institutional care costs are NHS/PSS this leads to a much larger ICER of £85,942. Thus, as with the Eisai model, the different payer's perspective impacts greatly upon the estimated ICERs.

Further adjustments were made to the PenTAG model for a moderate cohort by assuming the drug costs reported by the Lundbeck submission (£64.80 per month vs PenTAG assumption of £71.28 per month). This reduced the ICER modestly to £77,419 from £85,942.

Although adjustments to the cost estimates impact upon the ICERs obtained from the PenTAG model, the greatest difference between the models is the estimated delay to FTC/institutional care due to memantine. The AD progression model in the Lundbeck submission is based on ADAS-cog, ADCS-ADL and NPI. Thus an effect seen on all three scales will be likely to have a greater impact on the estimated time to FTC than an effect on just two scales as in the PenTAG model. The effectiveness estimates are shown in *Table 137*, and are of a similar magnitude for ADCS-ADL between the two models. Note that no estimates of ADAS-cog could be found from the included studies in Section 4 and were therefore not included as a sensitivity analysis in the PenTAG analyses. See Section 4.4 for further discussion of the clinical effectiveness from the Lundbeck submission.

TABLE 137 *Effectiveness estimates used in the Lundbeck and PenTAG models for memantine in a cohort of people with moderate AD*

Parameter	Lundbeck estimate	PenTAG estimate
ADAS-cog	-1.54	
MMSE		-0.7
ADCS-ADL	1.53	1.41
NPI	-1.34	

The different datasets on which disease progression was based between the Lundbeck and PenTAG models (LASER-AD vs the Wolstenholme study) demonstrates a particular source of structural uncertainty inherent in the decision modelling of treatments for people with Alzheimer's disease. Although the LASER-AD dataset is more recent and larger than the data from Oxfordshire in the Wolstenholme study, it has the disadvantage of containing many people who were already on AChEIs or memantine. The dataset from the study by Stern and colleagues[191] used in the SHTAC model offers a further option for predicting disease progression, and as seen in *Table 131* which provides different estimates of cost-effectiveness again.

8. Other factors relevant to the NHS

The care and treatment of people with Alzheimer's disease is complex and goes beyond the patient themselves to include carers to a degree not seen in many other conditions. The extra burden to the NHS, to social care services and the economy posed by the ill health of carers due to sub-optimal service provision for Alzheimer's patients is unknown, but must be considerable and growing. Unfortunately none of the trials included in the clinical effectiveness systematic review measured the effects of Alzheimer's disease on carers.

With respect to the economic evaluation many of the factors which would often be mentioned as "other factors" in this section, such as impact on carers, have already been highlighted in previous appraisals and directly considered in the modelling exercises. Such themes have been further pursued in the analysis in this report and are thus not mentioned here. Taking a wider societal perspective in the economic analyses is an issue which has been raised previously in relation to this topic. The reasons why such a broad perspective is not appropriate for the decisions made by NICE have already been clearly expressed and tested. To be consistent with this the main focus of analyses in this report has been from an NHS and Personal Social Service perspective. This is not to deny the value of taking a wider perspective in the context of other decisions outside NICE.

9. Discussion

9.1. Statement of principal findings

9.1.1. Aim

The remit for this report has been to up-date the evidence used to inform the last NICE guidance on donepezil, galantamine, rivastigmine and memantine for the treatment of Alzheimer's disease, particularly as laid out in the report by the Southampton Health Technology Assessments Centre (SHTAC). In general they considered evidence up to 2004, and this is the start date we have used for this report.

In this section we will not re-state the previous evidence, but assume that it will be read in the context of the previous evidence summaries and the decisions which flowed from them. Similarly the conclusions will focus on whether the new evidence on effectiveness and cost-effectiveness is likely to change the current guidance. A complete re-examination of all the available evidence from scratch was beyond the scope of this report.

9.1.2. Effectiveness review

In the previous assessment report in 2004, there was evidence for the effectiveness of donepezil, galantamine and rivastigmine on improving cognition, function, behaviour and global impact over the short term and evidence on the effectiveness of memantine was much more uncertain. Important gaps in the evidence were identified concerning long-term outcomes, impact on quality of life, carers and time to institutionalisation.

Overall we found that although more evidence has accumulated over the last six years, its impact on conclusions about effectiveness appears small. An enduring problem is that of trying to predict what will happen to people over the course of five years or more on the basis of six months or less information. The quality of many of the recent trials is a contributing factor to this; some good quality trials have been conducted but most of the new studies were of moderate to poor quality. A particular criticism is the use of LOCF and OC methods to account for missing data; these methods are inappropriate in a condition which naturally

declines to death and may lead to an overestimation of the treatment effect. Methods of randomisation and allocation concealment were frequently not reported.

In total, 17 new RCTs were included in the clinical effectiveness systematic review: there were 12 pairwise comparisons with placebo (donepezil 5, n=234; galantamine 3, n=1386; rivastigmine 3, n=1995 and memantine 1, n=350); four head-to-head studies and one combination therapy study (memantine added to AChEIs) were also found. The amount of evidence for these treatments has thus increased and has particularly consolidated the evidence on effectiveness of galantamine and rivastigmine relative to placebo. Evidence on the effectiveness of memantine does not appear to have been greatly strengthened. None of the gaps in evidence noted previously has been closed by new RCTs and no new evidence has emerged on differential effectiveness by sub-group, particularly disease severity and there is no evidence that these treatments increase longevity. Concerning comparative research, although there is one good quality new head-to-head trial, comparing donepezil and rivastigmine, showing results for rivastigmine on functional and global outcomes were significantly better than those for donepezil, more generally the case for one AChEI being more effective than another remains unconvincing. Our view, overall, is that these drugs should be treated as a class. The evidence about memantine hinged on two trials (one new and one from the previous review). The new study did not find any significant gain from memantine on any outcome. Although, pooling of these data with the previous review showed some inconsistent, partly positive, evidence on cognitive, functional and global outcomes.

In 2004 the assessment group found that donepezil improved cognitive and global outcomes, with increased benefit from higher doses, in some cases this benefit was maintained over a year. There was weaker evidence for a significant effect with functional and behavioural outcomes. The 2010 systematic review found five small poor quality studies which have added to the evidence base. They had a maximum of six months follow-up. All studies measured cognitive outcomes. A dose related beneficial effect was found at 10 mg/day. One study measured functional and global outcomes but it was of such poor quality the positive findings lack credibility.

We found an additional three variable quality RCTs of galantamine v. placebo to add to the evidence base of six studies included in 2004. The previous review found a dose-response relationship for cognitive, functional and global outcomes. In the two trials reporting

behavioural outcomes, one found a significant gain, the other did not. The studies included in our review all found significant benefit on cognitive outcomes; the results for functional and global outcomes were inconclusive, and no significantly positive gain was found for behavioural outcomes. However, when the results from these studies were pooled, significant gains for people taking galantamine were found for cognitive, functional and global outcomes.

The evidence for the effectiveness of rivastigmine in the previous review was varied; there was some evidence of benefit at 6-12mg/day with cognitive, functional and global outcomes, but no gain was reported on behavioural measures. Our update review found three more studies; one of these was of reasonable size and quality. Positive benefits from rivastigmine were found on cognitive, functional and global outcomes, but, as before, not on behavioural ones. The lower dose transdermal patch (9.5 mg/day) was shown to be as effective as the capsule (12 mg/day) but with fewer side effects.

There was some evidence, from a single study, in the previous review that memantine was more effective on cognitive and functional outcomes than placebo; although, as this study's results were not analysed by ITT, they may be unreliable. However, the new, poorer quality study, failed to show any benefit from memantine on any outcome measure. When the data were pooled, a significant benefit from memantine was found from global outcomes. It should be noted that these results are based on two moderate to poor quality trials and may be untrustworthy.

Three new head-to-head comparisons were found in addition to the three in the previous review. Only one of the new studies was large and of reasonable quality, this compared donepezil to rivastigmine. It measured cognitive, functional, behavioural and global outcomes, but only found statistically significant differences on functional and global outcomes, both favouring rivastigmine. This is in contrast to the much smaller and poorer quality studies found in the previous review, which showed no significant differences between the treatments. One new study and one previous study compared donepezil with galantamine; neither were good quality. The trial from the previous review found that donepezil had greater effects on cognitive and functional outcomes. The new study only looked at global outcomes and found no difference between the treatments. One very poor quality study, looking at behavioural outcomes, compared all three AChEIs; it found that rivastigmine was significantly better than donepezil or galantamine.

We also found one new, reasonably good, study comparing combined memantine with an AChE inhibitor against AChE inhibitor and placebo. This showed no significant advantage to combining these treatments. This contrasts with the results from the previous review which found significant benefits from combination therapy on cognitive, functional, behavioural and global outcomes. The reason for this difference in outcomes may be due to an underlying pharmacological interaction between galantamine and memantine - which neutralizes their respective effects - in the new trial, which used all three AChE inhibitors, whilst the existing trial only combined memantine with donepezil. The other difference between these studies is the lack of ITT analysis in the former one which may have led to more favourable results for combination therapy.

Mixed treatment comparison results varied depending on the outcome measure used. There was evidence for both donepezil and galantamine being probably the most effective treatment on cognitive outcomes. A similarly unclear picture for functional measures emerged with galantamine or rivastigmine possibly being equally effective. The amount of uncertainty in these results means it is impossible to say whether one AChE inhibitor is better than another at treating Alzheimer's disease.

9.1.2.1. Comparison with other systematic reviews

The findings of our systematic review comparing rivastigmine with placebo were similar to those of Birks and colleagues;[87] that rivastigmine confers benefit for those with mild to moderate Alzheimer's disease and that the benefit increases with increasing dose up to 12 mg/day, if the side effects can be tolerated. The transdermal patch, which confers similar clinical benefit to the capsule but with fewer adverse effects, may be a solution for some people who find that they cannot endure the capsules. Our findings differed slightly from those of the IQWiG (the German federal agency for assessing health technologies),[88] in that we did not find new evidence to support the assertion that galantamine could relieve psychological symptoms. This difference can be explained by their broader study design inclusion criteria. Otherwise, we agreed that the AChE inhibitors provided some help with the cognitive and functional symptoms from Alzheimer's disease. Hansen and colleagues,[95] who conducted a systematic review of functional outcomes from all the Alzheimer's drugs included in this review, found an overall benefit from treatment. This broadly agrees with our findings, although, the evidence for benefit from galantamine was inconsistent.

Three effectiveness reviews were submitted as part of manufacturer submissions in support of donepezil, galantamine and memantine. Full detail on the review method was only provided for the systematic review on donepezil. All reviews focused on RCTs although some non-RCT literature was also included. The comprehensiveness of the identification of this non-RCT literature was unclear. The new studies identified in the manufacturer submissions were consistent with those included in the PenTAG systematic review. The direction and size of effect relative to placebo on cognition, function, behaviour and global impact were consistent between the manufacturer submissions and the PenTAG systematic review. In the case of memantine the summary estimates of effect were more precise in the maufacturer's submission. Sub-group analyses for galantamine indicated that there was generally greater effectiveness for more severe AD. These analyses could not be done in the PenTAG systematic review. Evidence for the equal effectiveness of donepezil in mild AD relative to other severity groups argued against the presence of a sub-group effect. Sub-group analyses for memantine also did not show any difference in effectiveness by severity of AD, but did show a difference depending on the presence of Agitation/Aggression and/or Psychotic Symptoms (APS). Again these analyses could not be done in the PenTAG systematic review. Additional effects supported by non-RCT and observational data on duration of effectiveness, effects on carers, anti-psychotic use, institutionalisation and mortality were also claimed.

9.1.3. Economic evaluations

9.1.3.1. Initial estimates

The starting point for estimates of cost-effectiveness of donepezil, galantamine rivastigmine and memantine is complicated by the fact that the ICERs presented in the 2004 SHTAC report were considerably modified by discussions, debate and further work undertaken as part of the NICE appraisal process. The directly quoted cost per QALY gained in the NICE guidance document which underpinned the final decisions were:

- AChEIs for AD of moderate severity - £31,550 per QALY (the CQG quoted was specifically for donepezil)
- AChEIs for AD of mild severity - £55,000 to 58,000 per QALY, but with note that the true value was probably less than this, but not within the range normally considered cost-effective

- Memantine for severe AD – "above £53,000 per QALY"

9.1.3.2. Published economic evaluations

A systematic review of economic evaluations was conducted which identified 23 included studies published since 2004, over a third of which were only published as abstracts and could not be considered in depth. Of the remainder most addressed the costs and cost-effectiveness of either donepezil or memantine. Of these, the majority reapplied modelling approaches considered as part of the last guidance to the circumstances applying in other countries and were thus felt to add little to this update reconsidering cost-effectiveness in England and Wales. Enhanced modelling approaches were presented for both donepezil and memantine, but in both cases the publications closely mirrored the economic models submitted as part of the industry submissions, which we discuss in detail in the next section.

The included economic evaluations also provide some additional evidence on the impact on resource use and cost along-side trials. They provide support for the conclusion that use of donepezil or galantamine can be cost-saving in the short-term (6m to 1y). They conflict with the conclusion of the AD2000 study, the main economic evaluation alongside a trial included in the SHTAC report, which concluded that introduction of donepezil would increase costs over 2 years.

9.1.3.3. Industry submissions

Two companies offered models of cost-effectiveness: Eisai Ltd and Pfizer Ltd for donepezil and Lundbeck for memantine. Shire for galantamine made a submission focussing on effectiveness and emphasising issues concerning cost-effectiveness raised in the last appraisal; there was no submission for rivastigmine.

The model for donepezil has been described as a discrete event simulation model. This is a modelling approach which theoretically could overcome a number of challenges facing the assessment of the cost-effectiveness of drug treatments for AD, particularly dealing with multiple interdependent outcomes. However, the model does not employ a pure discrete event simulation approach and actually incorporates elements of individual sampling alongside some cohort modelling methods. The manufacturer's conclusion is that donepezil provides benefits at reduced costs relative to best supportive care, and is thus dominant, in

both mild and moderately severe AD, a conclusion which is robust to the sensitivity analyses conducted by the manufacturer. However, the review of the submitted model identified several areas where there was concern with respect to the quality of the inputted data or the validity of the model assumptions. Exploratory sensitivity analyses examining plausible alternative assumptions suggest that the cost-effectiveness could be at the margins of what would normally be considered cost-effective by NICE.

The model for memantine used a more traditional Markov approach with three states, pre full time care, full time care and death. It concludes that memantine provides benefits at reduced costs relative to best supportive care, and is thus dominant, in moderate and severe AD. Detailed appraisal again suggests that considerable caution is required in accepting this result with simple sensitivity analyses conducted by the report authors indicating ICERs which would not normally be considered cost-effective by NICE.

9.1.3.4. PenTAG cost–utility model

Despite modifications to overcome problems highlighted in the last appraisal, the results of the PenTAG model were not dissimilar to the results for the last TAR indicating that neither AChEIs nor memantine are cost-effective irrespective of the severity of AD being considered. This is attributable to failing to find cost-savings when the anti-AD treatments are employed, coupled with much smaller modelled estimates of health benefit relative to the manufacturers' submissions. It needs to be highlighted that the changes in effectiveness and cost underlying the ICERs are very small and that the results are highly uncertain and very sensitive to changes in several model assumptions and parameters.

In considering the strengths and weaknesses of the PenTAG model-based analyses, compared with the manufacturer and other models (see below), there should be no initial presumption that the model from the independent review group is somehow more valid or reliable than the others. Rather, in this complex disease area, the diversity of models - and resultant variation in the cost-effectiveness estimates - is partly a reflection of evident structural uncertainty regarding how to simulate this disease and its consequences, as well as differences in the rationales and context for developing each model. The PenTAG model, for example, has been developed in four to five months, with particular expectations to address some of the identified weaknesses of the previous model, and to be a single model capable of evaluating all the treatment comparators at different levels of disease severity

(both the AChEIs and memantine). The manufacturers, in contrast, have had a longer time period in which to develop their models, full access to their own trial data with which to inform them, and the more specific goal of evaluating the cost-effectiveness of their product.

For the AChEIs, the probabilistic sensitivity analyses suggested rivastigmine patches ($10cm^2$) were the most cost-effective of the AChEIs, but only with a probability of 17% of being the most cost-effective option at a willingness to pay of £30,000 per QALY (15% at a willingness to pay of £20,000 per QALY). Best supportive care was found to be the most cost-effective option with a probability of 57% of being cost-effective at a willingness to pay of £30,000 per QALY (62% at a willingness to pay of £20,000 per QALY). When compared to the next cheapest, non-dominated technology, the estimated deterministic ICER for rivastigmine patches compared to best supportive care was £61,100, and galantamine (16-24mg) was associated with an ICER of £157,800 per QALY compared to rivastigmine patches. Both donepezil (10mg) and rivastigmine capsules (9-12mg) were dominated. These ICERs should be interpreted with caution in light of the very small incremental costs and benefits and the considerable parameter and structural uncertainty in the PenTAG model.

For memantine in the treatment of moderate to severe AD, the probabilistic sensitivity analysis results estimated a probability of less than 4% that memantine was the most cost-effective option when compared to best supportive care at a willingness to pay of £30,000 per QALY (with a probability of 2.6% at a willingness to pay of £20,000 per QALY). The deterministic ICER was estimated to be £248,500 per QALY for a moderate to severe cohort. Although a great deal of parameter and structural uncertainty was also present in the cost-effectiveness analysis of memantine, none of the alternative assumptions assessed in the report lead to a positive net benefit for memantine compared to best supportive care at a willingness to pay of £30,000 per QALY.

Again it must be repeated that although the ICERs of all drugs are large relative to best supportive care, the incremental net benefits per patient are extremely small, given that the incremental costs and benefits are very small. This implies that funding all the drugs would reduce the total net benefit of the health service only by a very small amount.

9.2. Strengths and limitations of the systematic review of studies of effectiveness

The strengths of this systematic review are that is was conducted by an independent research team using the latest evidence.

There are a number of limitations:

- The length of follow up of the trials was a maximum of six months, which makes it very difficult to reliably extrapolate findings for years ahead.
- There has been a lack of evidence from the trials on key outcomes such as mortality, institutionalization, the impact on carer's time and the prescription of anti-psychotics.
- None of the trials conducted sub-group analyses based on disease severity, making us unable to comment on the effectiveness of treatments for mild, moderate or severe AD separately.
- Overall the quality of the trials was moderate to poor, with lack of reporting of key measures of trial quality, thus adding to the uncertainty of the results.
- The use of LOCF and OC methods for accounting for missing data may have overestimated the treatment benefit from the drugs.
- Some of the measures used in the trials are insensitive to change in Alzheimer's disease (ADAS-cog, MMSE). Therefore, the effects of treatment may have been underestimated in some cases.
- The searches were limited to the English language due to resource limitations, which may have led us to exclude important studies.

9.3. Strengths and limitations of the economic modelling by PenTAG

Although we believe we have made a number of improvements on the previous SHTAC-AHEAD model, and attempted to address some of the specific criticisms of the previous model (as detailed in Appendix 12), it should still be regarded as an exploratory model for

assessing the cost-effectiveness of drug treatments in this highly complex disease area. The main reasons for viewing the updated model and its outputs with such caution are:

- The underlying disease model captures just the two dimensions of cognitive status and functional status/ADL. Behavioural and psychological symptoms are not incorporated into the model, and therefore any treatment effects and quality of life impacts related to these symptoms will not be captured.

- The expression of treatment effectiveness, while based on a multivariate formula based on patient age, ADL status and cognitive status, is mainly based on predicting delays in time-to-institutionalization. While there is good evidence that this event/transition marks a key change in care costs, the evidence that it is also a key marker of decline in quality of life is uncertain.

- Although the model now incorporates more graduated declines in patient utility, and more graduated increases in NHS and PSS costs prior to institutionalization, assuming that all of these time-related cost and utility changes will be delayed by the same amount of time that institutionalization is delayed is a key assumption in the model (especially bearing in mind that many of the health care costs will not be related to Alzheimer's)

- The main database of individual patient data from the UK that the time-to-institutionalization model and key cost parameters are largely based upon is relatively old (1988-1999), small (n=92 with AD) and from a small part of the UK (Oxfordshire). Its generalisability to England and Wales in 2010 therefore has to be considered (see below).

Unlike the 2004 SHTAC analysis, utility benefits pre-institutionalization have been accounted for since utilities are based upon MMSE, and both costs and MMSE prior to institutionalization are conditional on time until institutionalization. However, as with the previous model, basing the simple structure of the model around the two main stages of living in the community (i.e. at home), or living in a nursing or residential home (or long-term hospitalisation), means estimating the benefits of drug treatments for those already in residential care is problematic. This is a more considerable weakness of this modeling approach for evaluating the cost-effectiveness of memantine.

In attempting to overcome a criticism of the SHTAC model where AD progression was based on US data, AD progression in the PenTAG model is based on UK individual patient data. However, the generalisability of this data should be questioned for a number of reasons: (i) the data are from just 92 individuals, (ii) it is collected from the Oxfordshire area only, and (iii) these data were collected between 1988/9 and 1999. Not only are these data used to inform AD progression, they are also used as a basis for the NHS/PSS costs of care (in the community and in institutions). This has an advantage in one respect since there is no need to incorporate an additional source of evidence, with its own uncertainties, into the model. However if the data from Wolstenholme and colleagues cannot be generalised to the situation in England and Wales in 2010, it is likely the model will not be generalisable either, even though few options were available as the basis for predicting disease progression. In addition to considering the US data used in the SHTAC model, RCT data were considered but felt not to be ideal due to the restricted populations from inclusion/exclusion criteria. The available UK epidemiological evidence was either from Wolstenholme and colleagues or a longitudinal cohort study where many participants were receiving AChEI and/or memantine treatment (i.e. the LASER-AD study).

The incorporation of the full treatment effect at six months is artificial. It is more likely that improvements due to treatment are gradual. It is also assumed in the PenTAG model that treatment benefits remain after treatment has ceased. This assumption is also likely to be unrealistic, but is favourable to the active treatments. Furthermore, the treatment effects incorporated into the PenTAG model are absolute effects. There has been no accounting for differential effects for baseline severity, but there was some, albeit exploratory, evidence of an association between baseline MMSE and functional outcomes identified in Section 4 (see Appendix 7).

A further limitation relates to effectiveness data availability. No relevant ADL data for donepezil and no relevant MMSE for galantamine at 21-26 weeks were identified from the clinical effectiveness review. Thus, it was assumed that this was a lack of evidence for an effect, rather than lack of effect and a class effect was assumed (i.e. the effectiveness was assumed to be the same as the other AChEIs).

9.4. Strengths and limitations of the economic modelling in the Eisai/Pfizer submission

A strength of the Esai/Pfizer model is that it is able to track changes on cognitive status, functional status (using both ADL and IADL), and behavioural and psychological symptoms. However, there were a number of concerns with the appropriateness of the data used to predict progression on each of these scales and the possibility of double counting treatment effects, since changes on one scale were used as an independent term to predict progression on the other scales. The electronic version of the model contained many features which were not used in the submission and the presence of these redundant features reduced the transparency of the model making it very difficult to review. Whilst some errors were identified, we cannot be entirely confident that no other errors remain unidentified. The review as a whole cannot be considered as an endorsement of the validity of the model. We were unable to explain some features of the behaviour of the model and therefore retain a degree of caution about its functioning.

The most significant weakness with the model is that the data used in the model to relate cognitive function (MMSE) to the probability of institutionalisation appears to have been derived from a study which only included institutionalised patients and insufficient details are provided to explain how the data used in the model could have been derived from this study. This is a significant weakness as it is a major driver of cost-effectiveness.

9.5. Uncertainties

There continue to be many uncertainties, indeed it is likely that the nature and extent of these uncertainties is similar to those operating when the last TAR was compiled. The most influential of these are:

- Effect of anti-AD drugs in the longer term on any outcome, especially beyond one year.
- Effect of anti-AD drugs on outcomes beyond cognition, function, behaviour and global impact, particularly quality of life, impact on carers, effect on admission to full time care and impact on resource use.
- Whether the effects vary substantially by sub-group, particularly severity of AD.

- Which modelling approach delivers the most accurate assessment of cost-effectiveness in AD.
- Whether the future cost of the anti-AD drugs will be affected by the entry of generic formulations[217]

10. Conclusions

The additional clinical effectiveness evidence identified in this up-date systematic review continues to suggest that there is clinical benefit from the AChEIs in alleviating symptoms and controlling disease progression in AD. However, there is only randomised evidence for this up to six months. Although there is also new evidence on the effectiveness of memantine, but it remains less supportive of this drug's use.

While there remains considerable debate about the magnitude of the effect of AChEIs on cognition, function, behaviour and global impact, there is very little, if any, disagreement that the effects are present.

Conclusions concerning cost-effectiveness are however no clearer. This arises from uncertainty about the most appropriate modelling approach, compounded by uncertainty about all model parameters. Although we can explain some of the large differences in the cost-effectiveness estimates between the industry submissions for donepezil and memantine and the PenTAG model, these cannot be completely accounted for.

Whatever the final judgment about the most likely true ICER values, it must be recognized that the estimates are based on very small incremental benefits and costs.

10.1. Implications for service provision

These are not clear and will ultimately rest on the interpretation of the new evidence from a variety of sources, including this report, in the forthcoming NICE appraisal on this topic.

10.2. Suggested research priorities

New research in the following areas could reduce the uncertainty noted:

- Good quality longer term RCTs (following CONSORT) to include mortality, time to institutionalization and HR QOL as outcomes and sufficiently powered for subgroup analysis by disease severity, response to treatment, behavioural disturbance and comorbidities. We have identified that a limited number of major RCTs addressing

relevant issues such as management when patients fail to respond to AChEIs are already in progress (DOMINO-AD).

- Such good quality trials should aim to use the same standardized measures of cognitive status, functional status/ADL, and behavioural/psychiatric symptoms.
- Systematic reviews of non-RCT evidence on the impact of anti-AD treatments on resource use, institutionalisation and mortality.
- Further independent comparison of different methodological approaches to modelling the cost-effectiveness of anti-AD treatments.
- Research into cognitive measures that are sensitive to change in dementia.
- Studies should measure HRQoL with the DEMQOL which has been validated for use with dementia patients rather than the EQ-5D which has not. Work is needed to derive utility values from the DEMQOL or to map it onto the EQ-5D or HUI 2/3

In addition this report highlights some wider methodological issues which would benefit from further investigation:

- Research into more valid ways of accounting for missing data than LOCF and OC particularly in degenerative diseases like AD.

References

(1) National Institute of Health and Clinical Excellence. Donepezil, galantamine, rivastigmine (review) and memantine for the treatment of Alzheimer's disease (amended). TA 111. 2009. London, NICE. NICE technology appraisal guidance.

(2) Loveman E, Green C, Kirby J, Takeda A, Picot J, Payne E, Clegg A. The clinical and cost-effectiveness of donepezil, rivastigmine, galantamine and memantine for Alzheimer's disease. TA 111. 2004. Southampton, University of Southampton.

(3) McKhann G, Drachman D, Folstein M, Katzman R, Price D, Stadlan EM. Clinical diagnosis of Alzheimer's disease: report of the NINCDS-ADRDA Work Group under the auspices of Department of Health and Human Services Task Force on Alzheimer's Disease. Neurology 1984; 34(7):939-944.

(4) National Collaborating Centre for Mental Health. Dementia: a NICE-SCIE Guideline on supporting people with dementia and their carers in health and social care. 42. 2007. Leicester, The British Psychological Society and Gaskell. National Clinical Practice Guideline.

(5) National Audit Office. Improving services and support for people with dementia. 2007. London, The Stationary Office.

(6) Knapp M, Prince M, Albanese E, Banerjee S, Dhanasiri S, Fernandez J, Ferri C, Knapp M, McCrone P, Prince M, Snell T, Stewart R. Dementia UK: The full report. 2007. London, Alzheimer's Society.

(7) Luengo-Fernandez R, Leal J, Gray A. Dementia 2010: the economic burden of dementia and associated research funding in the United. 2010. Cambridge, Alzheimer's Research Trust.

(8) Yip AG, Brayne C, Matthews FE. Risk factors for incident dementia in England and Wales: The Medical Research Council Cognitive Function and Ageing Study. A population-based nested case-control study. Age Ageing 2006; 35(2):154-160.

(9) Matthews F, Brayne C. The incidence of dementia in England and Wales: findings from the five identical sites of the MRC CFA Study. PLoS Med 2005; 2(8):e193.

(10) Alzheimer's Association., Alzheimer's Association. 2009 Alzheimer's disease facts and figures. Alzheimer's & Dementia 2009; 5(3):234-270.

(11) Hardy J. New insights into the genetics of Alzheimer's disease. Ann Med 1996; 28(3):255-258.

(12) Cruts M, van Duijn CM, Backhovens H, Van den BM, Wehnert A, Serneels S, Sherrington R, Hutton M, Hardy J, St George-Hyslop PH, Hofman A, van Broeckhoven C. Estimation of the genetic contribution of presenilin-1 and -2 mutations in a population-based study of presenile Alzheimer disease. Hum Mol Genet 1998; 7(1):43-51.

(13) Schellenberg G, Anderson L, O'dahl S. APP717, APP693 and PRIP gene mutations are rare in Alzheimer's disease. American Journal of Human Genetics 1991; 49:511-517.

(14) Takashi ASAD. Prevention of Alzheimer's disease: Putative nutritive factors. Psychogeriatrics 2007; 7(3):125-131.

(15) Rolland YM, Pillard FM, Klapouszczak AM, Reynish EM, Thomas DM, Andrieu SM, Riviere D, Vellas B. Exercise Program for Nursing Home Residents with Alzheimer's Disease: A 1-Year Randomized, Controlled Trial. Journal of the American Geriatrics Society 2007; 55(2):158-165.

(16) Fratiglioni L, Paillard-Borg S, Winblad B. An active and socially integrated lifestyle in late life might protect against dementia. The Lancet Neurology 2004; 3(6):343-353.

(17) Fitzpatrick AL, Kuller LH, Lopez OL, Kawas CH, Jagust W. Survival following dementia onset: Alzheimer's disease and vascular dementia. Journal of the Neurological Sciences 2005; 229-230:43-49.

(18) Oxford Textbook of Medicine. Fourth ed. Oxford: Oxford University Press; 2003.

(19) Backman L, Jones S, Berger AK, Laukka EJ, Small BJ. Multiple cognitive deficits during the transition to Alzheimer's disease. J Intern Med 2004; 256(3):195-204.

(20) Arnaiz E, Almkvist O. Neuropsychological features of mild cognitive impairment and preclinical Alzheimer's disease. Acta Neurol Scand Suppl 2003; 179:34-41.

(21) Ritchie K, Lovestone S. The dementias. Lancet 2002; 360(9347):1759-1766.

(22) Brodaty H, Hadzi-Pavlovic D. Psychosocial effects on carers of living with persons with dementia. Aust N Z J Psychiatry 1990; 24(3):351-361.

(23) Carlesimo GA, Oscar-Berman M. Memory deficits in Alzheimer's patients: a comprehensive review. Neuropsychol Rev 1992; 3(2):119-169.

(24) Frank EM. Effect of Alzheimer's disease on communication function. J S C Med Assoc 1994; 90(9):417-423.

(25) Wimo A, Gustafsson L, Mattson B. Predictive validity of factors influencing the institutionalization of elderly people with psycho-geriatric disorders. Scand J Prim Health Care 1992; 10(3):185-191.

(26) Drachman DA, O'Donnell BF, Lew RA, Swearer JM. The prognosis in Alzheimer's disease. 'How far' rather than 'how fast' best predicts the course. Arch Neurol 1990; 47(8):851-856.

(27) Brodaty H, McGilchrist C, Harris L, Peters KE. Time until institutionalization and death in patients with dementia. Role of caregiver training and risk factors. Arch Neurol 1993; 50(6):643-650.

(28) Thompson C, Spilsbury K, Hall J, Birks Y, Barnes C, Adamson J. Systematic review of information and support interventions for caregivers of people with dementia. BMC Geriatrics 2007; 7(1):18.

(29) Aguglia E, Onor. Stress in the caregivers of Alzheimer's patients: An experimental investigation in Italy. American Journal of Alzheimer's Disease and other Dementias 2004; .19(4).

(30) Holland JM, Currier JM, Gallagher-Thompson D. Outcomes from the Resources for Enhancing Alzheimer's Caregiver Health (REACH) Program for Bereaved Caregivers. Psychology & Aging 2009; 24(1):190-202.

(31) Garity JE. Caring for a Family Member with Alzheimer's Disease: Coping with Caregiver Burden Post-Nursing Home Placement. Journal of Gerontological Nursing 2006; 32(6):39-48.

(32) Vitaliano PP, Echeverria D, Yi J, Phillips PEM, Young H, Siegler IC. Psychophysiological Mediators of Caregiver Stress and Differential Cognitive Decline. Psychology & Aging 2005; 20(3):402-411.

(33) Bianchetti A, Scuratti A, Zanetti O, Binetti G, Frisoni GB, Magni E, Trabucchi M. Predictors of mortality and institutionalization in Alzheimer disease patients 1 year after discharge from an Alzheimer dementia unit. Dementia 1995; 6(2):108-112.

(34) Donaldson C, Tarrier N, Burns A. The impact of the symptoms of dementia on caregivers. Br J Psychiatry 1997; 170:62-68.

(35) Adams KB, Sanders SA. Alzheimer's caregiver differences in experience of loss, grief reactions and depressive symptoms across stage of disease: A mixed-method analysis. Dementia 2004; 3(2):195-210.

(36) Schneider J, Murray J, Banerjee S, Mann A. EUROCARE: a cross-national study of co-resident spouse carers for people with Alzheimer's disease: I--Factors associated with carer burden. Int J Geriatr Psychiatry 1999; 14(8):651-661.

(37) Mittelman MS, Haley WE, Clay OJM, Roth DLP. Improving caregiver well-being delays nursing home placement of patients with Alzheimer disease. Neurology 2006; 67(9):1592-1599.

(38) Banerjee S, Murray J, Foley B, Atkins L, Schneider J, Mann A. Predictors of institutionalisation in people with dementia. Journal of Neurology, Neurosurgery & Psychiatry 2003; 74(9):1315-1316.

(39) de Vugt ME, Stevens F, Aalten P, Lousberg R, Jaspers N, Verhey FRJ. A prospective study of the effects of behavioral symptoms on the institutionalization of patients with dementia. International Psychogeriatrics 2005; 17(04):577-589.

(40) Wolfson C, Moride Y, Perrault A. Drug treatments for Alzheimer's disease. II: a review of outcome measures in clinical trials. 2000. Ottawa, Canadian Coordinating Office for Health Technology Assessment.

(41) Wolfson C, Oremus M, Shukla V, Momoli F, Demers L, Perrault A, Moride Y. Donepezil and rivastigmine in the treatment of Alzheimer's disease: a best-evidence synthesis of the published data on their efficacy and cost effectiveness. Clinical Therapeutics 2002; 24(6):862-886.

(42) Folstein MF, Folstein SE, McHugh PR. "Mini-mental state" : A practical method for grading the cognitive state of patients for the clinician. Journal of Psychiatric Research 1975; 12(3):189-198.

(43) Mohs RC, Rosen WG, Davis KL. The Alzheimer's disease assessment scale: an instrument for assessing treatment efficacy. Psychopharmacol Bull 1983; 19(3):448-450.

(44) Roselli F, Tartaglione B, Federico F, Lepore V, Defazio G, Livrea P, Roselli F, Tartaglione B. Rate of MMSE score change in Alzheimer's disease: influence of education and vascular risk factors. Clinical Neurology & Neurosurgery 2009; 111(4):327-330.

(45) Tombaugh TN. Test-retest reliable coefficients and 5-year change scores for the MMSE and 3MS. Archives of Clinical Neuropsychology 2005; 20(4):485-503.

(46) Irizarry MC, Webb DJ, Bains C, Barrett SJ, Lai RY, Laroche JP, Hosford D, Maher-Edwards F, Weil JG. Predictors of placebo group decline in the Alzheimer's disease Assessment Scale-cognitive subscale (ADAS-Cog) in 24 week clinical trials of Alzheimer's disease. Journal of Alzheimer's Disease 2008; 14(3):301-311.

(47) Cummings JL, Mega M, Gray K, Rosenberg-Thompson S, Carusi DA, Gornbein J. The Neuropsychiatric Inventory: comprehensive assessment of psychopathology in dementia. Neurology 1994; 44(12):2308-2314.

(48) Hughes CP, Berg L, Danziger WL, Coben LA, Martin RL. A new clinical scale for the staging of dementia. Br J Psychiatry 1982; 140:566-572.

(49) Boothby H, Mann AH, Barker A. Factors determining inter-rater agreement with rating global change in dementia: CIBIC-plus. Int J Geriatr Psychiatry 1995; 10:1037-1045.

(50) Claus JJ, Teunisse S, Walstra GJM, van Gool WA. Determinants of Global Clinical Change Assessment in Patients with Early Alzheimer's Disease. Dement Geriatr Cogn Discord 1998;9:157-163.

(51) Gauthier S, Bodick N, Erzigkeit E, Feldman H, Geldmacher DS, Huff J, Mohs R, Orgogozo JM, Rogers S. Activities of daily living as an outcome measure in clinical trials of dementia drugs. Position paper from the International Working Group on Harmonization of Dementia Drug Guidelines. Alzheimer Dis Assoc Disord 1997; 11 Suppl 3:6-7.

(52) Nygard L. Assessing ADL/IADL in persons with dementia. In: Wimo A, Jonnson B, Karlsson G, Winblad B, editors. Health Economics of Dementia. New York: John Wiley and Sons; 1998. 371-388.

(53) McDowell I, Newell C. Measuring Health: a guide to rating scales and questionnaires. 2 ed. New York: Oxford University Press; 1996.

(54) Smith SC, Lampling DL, Banerjee S, Harwood R, Foley B, Smith P, Cook JC, Murray J, Prince M, Levin E, Mann A, Knapp M. Measurement of health-related quality of life for people with dementia: development of a new instrument (DEMQOL) and an evaluation of current methodology. [9 (10)]. 2005. Health Technology Assessment.

(55) Blau TH. Quality of life, social indicators and criteria of change. Prof Psychol 1977; 8:464-473.

(56) Department of Health. Living well with dementia: A National Dementia Strategy. 2009. London, Department of Health.

(57) Community care for frail older people: analysis using the 1998/99 General Household Survey. Quality in Later life: Rights, Rhetoric and Reality; Stirling, Scotland: University of Stirling; 2001.

(58) Audit Commission. Support for carers of older people. 2004. London, Audit Commission.

(59) Comas-Herrera A, Wittenberg R, Pickard L, Knapp M. Cognitive impairment in older people: future demand for long-term care services and the associated costs. Int J Geriatr Psychiatry 2007; 22(10):1037-1045.

(60) National Assembly for Wales. Caring about Carers: A Strategy for Carers in Wales. Implementation Plan. 2000. Cardiff, National Assembly for Wales.

(61) Secretary of State for Health. The NHS Plan. Cm 4818. 2000. London, The Stationary Office.

(62) Morris RG, Morris LW, Britton PG. Factors affecting the emotional wellbeing of the caregivers of dementia sufferers. Br J Psychiatry 1988; 153:147-156.

(63) Livingston G, Manela M, Katona C. Depression and other psychiatric morbidity in carers of elderly people living at home. BMJ (Clinical research ed) 1996; 312(7024):153-156.

(64) Medical Research Council. Cognitive function and dementia in six areas of England and Wales: the distribution of MMSE and prevalence of GMS organicity level in the MRC CFA Study. Psychological Medicine 1998; 28:319-335.

(65) Evandrou M. Employment and care, paid and unpaid work: the socio economic position of informal carers in Britain. In: Phillips J, editor. Working Carers. Aldershot: Avebury; 1995.

(66) Roth M, Tym E, Mountjoy CQ, Huppert FA, Hendrie H, Verma S et al. CAMDEX. A standardised instrument for the diagnosis of mental disorder in the elderly with special reference to the early detection of dementia. Br J Psychiatry 1986; 149:698-709.

(67) Department of Health, Care Services Improvement Partnership. Everybody's Business. Integrated mental health service for older adults: a service development guide. 2005. London, Department of Health.

(68) Audit Commission. Forget-Me-Not 2002: Mental Health Services for Older People. 2002. London, Audit Commission.

(69) Department of Health. National Service Framework for Older People. 2001. London, Department of Health.

(70) Audit Commission. Forget-Me-Not: Mental Health Services for Older People. 2000. London, The Audit Commission.

(71) NHS Centre for Reviews and Dissemination. Systematic Reviews: CRD's guidance for undertaking reviews in healthcare. York: CRD University of York; 2009.

(72) Moher D, Liberati A, Tetzlaff J, Altman DG, for the PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. BMJ (Clinical research ed) 2009; 339(jul21_1):b2535.

(73) Moher D, Cook DJ, Eastwood S, Olkin I, Rennie D, Stroup DF. Improving the quality of reports of meta-analyses of randomised controlled trials: the QUOROM statement. The Lancet 1999; 354(9193):1896-1900.

(74) DerSimonian R, Laird N. Meta-analysis in clinical trials. Control Clin Trials 1986; 7(3):177-188.

(75) Cochran WG. The Combination of Estimates from Different Experiments. Biometrics 1954; 10(1):101-129.

(76) Higgins JPT, Thompson SG. Quantifying heterogeneity in a meta-analysis. Statist Med 2002; 21(11):1539-1558.

(77) Higgins JPT, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. BMJ (Clinical research ed) 2003; 327:557-560.

(78) Egger M, Smith GD, Schneider M, Minder C. Bias in meta-analysis detected by a simple, graphical test. BMJ (Clinical research ed) 1997; 315(7109):629-634.

(79) The Cochrane Collaboration. Cochrane Handbook for Systematic Reviews of Interventions Version 5.0.2 [updated September 2009]. 2009.

(80) Molnar FJ, Hutton B, Fergusson D. Does analysis using "last observation carried forward" introduce bias in dementia research? Canadian Medical Association Journal 2008; 179(8):751-753.

(81) Thompson SG, Sharp SJ. Explaining heterogeneity in meta-analysis: a comparison of methods. Statistics in Medicine 1999; 18(20):2693-2708.

(82) Egger M, Smith GD, Altman DG. Systematic reviews in health care: meta-analysis in context. 2nd ed. London: BMJ Books; 2001.

(83) Ades AE. A chain of evidence with mixed comparisons: models for multi-parameter synthesis and consistency of evidence.(0277-6715 (Print)).

(84) Caldwell DM FAU, des AE FAU, Higgins JP. Simultaneous comparison of multiple treatments: combining direct and indirect evidence.(1468-5833 (Electronic)).

(85) Lu GF, Ades AE. Combination of direct and indirect evidence in mixed treatment comparisons.(0277-6715 (Print)).

(86) Higgins JP FAU, Whitehead A. Borrowing strength from external trials in a meta-analysis.(0277-6715 (Print)).

(87) Birks J, Evans JG, Iakovidou V, Tsolaki M. Rivastigmine for Alzheimer's disease. Cochrane Database of Systematic Reviews 2009; (2)(CD001191).

(88) IQWiG. Cholinesterase inhibitors in Alzheimer's disease. 2007. Cologne, Institut fuer Qualitaet und Wirtschaftlichkeit im Gesundheitswesen (IQWiG).

(89) Brodaty H, Corey-Bloom J, Potocnik FCV, Truyen L, Gold M, Damaraju CRV. Galantamine prolonged-release formulation in the treatment of mild to moderate Alzheimer's disease. Dementia and Geriatric Cognitive Disorders 2005; 20(2-3):120-132.

(90) Rockwood K, Fay S, Song X, MacKnight C, Gorman M. Attainment of treatment goals by people with Alzheimer's disease receiving galantamine: a randomized controlled trial. CMAJ Canadian Medical Association Journal 2006; 174(8):1099-1105.

(91) Bullock R. Treatment of behavioural and psychiatric symptoms in dementia: implications of recent safety warnings. Current Medical Research and Opinion 2005; 21(1):1-10.

(92) Cumbo E. Differential effects or rivastigmine, galantamine and donepezil on behavioral and psychological symptoms in patients with Alzheimer's disease: 18-month, randomized, open-label trial. Primary Care and Community Psychiatry 2005; 10(3):95-102.

(93) Raina P, Santaguida P, Ismaila A, Patterson C, Cowan D, Levine M, Booker L, Oremus M. Effectiveness of cholinesterase inhibitors and memantine for treating dementia: Evidence review for a clinical practice guideline. Annals of Internal Medicine 2008; 148(5):379-397.

(94) 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 2004; 17(1-2):29-34.

(95) Hansen RA, Gartlehner G, Lohr KN, Kaufer DI, Hansen RA, Gartlehner G, Lohr KN, Kaufer DI. Functional outcomes of drug treatment in Alzheimer's disease: A systematic review and meta-analysis. Drugs & Aging 2007; 24(2):155-167.

(96) Bentham P, Gray R, Raftery J, Hills R, Sellwood E, Courtney C, Farrell D, Hardyman W, Crome P, Edwards S, Lendon C, Lynch L. Long-term donepezil treatment in 565 patients with Alzheimer's disease (AD2000): randomised double-blind trial. Lancet 2004; 363(9427):2105-2115.

(97) Burns A, Rossor M, Hecker J, Gauthier S, Petit H, Moeller HJ, Rogers SL, Friedhooff LT. The effects of donepezil in Alzheimer's disease: results from a multinational trial. Dementia and Geriatric Cognitive Disorders 1999; 10(3):237-244.

(98) Gauthier S, Feldman H, Hecker J, Vellas B, Emir B, Subbiah P, Donepezil M. Functional, cognitive and behavioral effects of donepezil in patients with moderate Alzheimer's disease. Current Medical Research and Opinion 2002; 18(6):347.

(99) Greenberg SM, Tennis MK, Brown LB, Gomez-Isla T, Hayden DL, Schoenfeld DA, Walsh KL, Corwin C, Daffner KR, Friedman P. Donepezil Therapy in Clinical Practice: A Randomized Crossover Study. Arch Neurol 2000; 57(1):94-99.

(100) Holmes C, Wilkinson D, Dean C, Vethanayagam S, Olivieri S, Langley A, Pandita-Gunawardena ND, Hogg F, Clare C, Damms J. The efficacy of donepezil in the

treatment of neuropsychiatric symptoms in Alzheimer disease. Neurology 2004; 63(2):214-219.

(101) Homma A, Takeda M, Imai Y, Udaka F, Hasegawa K, Kameyama M, Nishimura T. Clinical efficacy and safety of donepezil on cognitive and global function in patients with Alzheimer's disease. A 24-week, multicenter, double-blind, placebo-controlled study in Japan. E2020 Study Group. Dementia and Geriatric Cognitive Disorders 11(6):299.

(102) Krishnan K, Charles HC, Doraiswamy PM, Mintzer J, Weisler R, Yu X, Randomized, placebo-controlled trial of the effects of donepezil on neuronal markers and hippocampal volumes in Alzheimer's disease. American Journal of Psychiatry 2003; 160(11).

(103) Mohs RC, Doody RS, Morris JC, Ieni JR, Rogers SL, Perdomo CA, Pratt RD. A 1-year, placebo-controlled preservation of function survival study of donepezil in AD patients. Neurology 2001; 57(3):481.

(104) Johannsen P, Barcikowska M, Dautzenberg P, Hampel H, Hasselbalch S, Sakka P, Tilker H, Tury F, Qvitzau S, Richardson S, Xu Y, Schwam E, Schindler R. Donepezil-treated Alzheimer's disease patients with apparent initial cognitive decline demonstrate significant benefits when therapy is continued: results from a randomised, placebo-controlled trial. International Psychogeriatrics 2003; 15: 106-107.

(105) Johannsen P, Salmon E, Hampel H, Xu Y, Richardson S, Qvitzau S, Schindler R. Assessing therapeutic efficacy in a progressive disease: A study of donepezil in Alzheimer's disease. CNS Drugs 2006; 20(4):311-325.

(106) Rogers SL, Farlow MR, Doody RS, Mohs R, Friedhoff LT. A 24-week, double-blind, placebo-controlled trial of donepezil in patients with Alzheimer's disease. Neurology 1998; 50(1):136.

(107) Rogers SL, Doody RS, Mohs RC, Friedhoff LT. Donepezil improves cognition and global function in Alzheimer disease: a 15-week, double-blind, placebo-controlled study. Archives of Internal Medicine 1998; 158(9):1021.

(108) Rogers SL, Friedhoff LT. The efficacy and safety of donepezil in patients with Alzheimer's disease: results of a US multicentre, randomized, double-blind, placebo-controlled trial. Dementia 1996; 7:293-303.

(109) Seltzer B, Zolnouni P, Nunez M, Goldman R, Kumar D, Leni J, Richardson S. Efficacy of donepezil in early-stage Alzheimer disease: A randomized placebo-controlled trial. Arch Neurol 2004; 61(12):1852-1856.

(110) Winblad B, Engedal K, Soininen H, Verhey F, Waldemar G, Wimo A, Wetterholm AL, Zhang R, Haglund A, Subbiah P. A 1-year, randomized, placebo-controlled study of donepezil in patients with mild to moderate AD. Neurology 2001; 57(3):489.

(111) Wimo A, Winblad B, Engedal K, Soininen H, Verhey F, Waldemar G, Wetterholm AL, Mastey V, Haglund A, Zhang R. An economic evaluation of donepezil in mild to moderate Alzheimer's disease: results of a 1-year, double-blind, randomized trial. Dementia and Geriatric Cognitive Disorders 2003; 15(1):44.

(112) Mazza M, Capuano A, Bria P, Mazza S. Ginkgo biloba and donepezil: A comparison in the treatment of Alzheimer's dementia in a randomized placebo-controlled double-blind study. European Journal of Neurology 2006; 13(9):981-985.

(113) Dos Santos Moraes WA, Poyares DR, Guilleminault C, Ramos LR, Ferreira Bertolucci PH, Tufik S. The effect of donepezil on sleep and REM sleep EEG in patients with Alzheimer disease: A double-blind placebo-controlled study. Sleep 2006; 29(2):199-205.

(114) Moraes W, Poyares D, Sukys-Claudino L, Guilleminault C, Tufik S. Donepezil improves obstructive sleep apnea in Alzheimer disease: A double-blind, placebo-controlled study. Chest 2008; 133(3):677-683.

(115) Peng DT, Xu XH, Wang LN. Efficiency and safety assessment of donepezil for treating mild and moderate Alzheimer disease. Chinese Journal of Clinical Rehabilitation 2005; 9(13):170-172.

(116) Winstein CJ, Bentzen KR, Boyd L, Schneider LS. Does the cholinesterase inhibitor, donepezil, benefit both declarative and non-declarative processes in mild to moderate Alzheimer's disease? Current Alzheimer Research 2007; 4(3):273-276.

(117) Raskind MA, Peskind ER, Wessel T, Yuan W. Galantamine in AD: a 6-month randomized, placebo-controlled trial with a 6-month extension. Neurology 2000; 54(12):2261.

(118) Rockwood K, Mintzer J, Truyen L, Wessel T, Wilkinson D. Effects of a flexible galantamine dose in Alzheimer's disease: a randomised, controlled trial. British Medical Journal 2001; 71(5):589.

(119) Tariot PN, Solomon PR, Morris JC, Kershaw P, Lilienfeld S, Ding C. A 5-month, randomized, placebo-controlled trial of galantamine in AD. The Galantamine USA-10 Study Group. Neurology 2000; 54(12):2269.

(120) Cummings JL, Schneider L, Tariot PN, Kershaw PR, Yuan W. Reduction of behavioral disturbances and caregiver distress by galantamine in patients with Alzheimer's disease. American Journal of Psychiatry 2004; 161(3):532-538.

(121) Wilcock GK, Lilienfeld S, Gaens E. Efficacy and safety of galantamine in patients with mild to moderate Alzheimer's disease: multicentre randomised controlled trial. British Medical Journal 2000; 321(7274):1445.

(122) Wilkinson D, Murray J. Galantamine: a randomized, double-blind, dose comparison in patients with Alzheimer's disease. International Journal of Geriatric Psychiatry 2001; 16(9):852.

(123) Wilkinson D, Lilienfeld S, Truyen L. Galantamine improves activities of daily living in patients with Alzheimer's disease: a 3 month placebo-controlled study. Poster: Sixth International Stockholm/Springfield Symposium on Advances in Alzheimer Therapy. 2000.

(124) 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. The Lancet 2002; 359(9314):1283-1290.

(125) Rockwood K, Fay S, Song X, MacKnight C, Gorman M. Attainment of treatment goals by people with Alzheimer's disease receiving galantamine: A randomized controlled trial. Canadian Medical Association Journal 2006; 174(8):1099-1105.

(126) Feldman HH, Van Baelen B, Kavanagh S. Effects of galantamine on activities of daily living in Alzheimer's disease: Evidence from six randomised double-blind placebo-controlled trials. Research and Practice in Alzheimer's Disease 2005; 10:234-238.

(127) Rockwood K, Fay S, Jarrett P, Asp E. Effect of galantamine on verbal repetition in AD: A secondary analysis of the VISTA trial. Neurology 2007; 68(14):1116-1121.

(128) Agid Y, Dubois B, Anand R, Gharabawi G. Efficacy and tolerability of rivastigmine in patients with dementia of the Alzheimer type. Current Therapeutic Research 1998; 59(12):837-845.

(129) Corey-Bloom J, Anand R, Veach J. A randomized trial evaluating the efficacy and safety of ENA 713(rivastigmine tartrate), a new acetylcholinesterase inhibitor, in patients with mild to moderately severe Alzheimer's disease. International Journal of Geriatric Psychopharmacology 1998; 1(2):55-65.

(130) Forette F, Anand R, Gharabawi G. A phase II study in patients with Alzheimer's disease to assess the preliminary efficacy and maximum tolerated dose of rivastigmine (Exelon [R]). European Journal of Neurology 1999; 6(4):423-429.

(131) Rosler M, Anand R, Cicin-Sain A, Gauthier S, Agid Y, Dal-Bianco P, Stalhelin HB, Hartman R, Gharabawi M, Bayer T. Efficacy and safety of rivastigmine in patients with Alzheimer's disease: international randomised controlled trial. Commentary: Another piece of the Alzheimer's jigsaw. British Medical Journal 1999; 318(7184):633.

(132) Feldman HH, Lane R. Rivastigmine: A placebo controlled trial of twice daily and three times daily regimens in patients with Alzheimer's disease. Journal of Neurology, Neurosurgery and Psychiatry 2007; 78(10):1056-1063.

(133) Mowla A, Mosavinasab M, Haghshenas H, Haghighi AB. Does serotonin augmentation have any effect on cognition and activities of daily living in Alzheimer's dementia? A double-blind, placebo-controlled clinical trial. Journal of Clinical Psychopharmacology 2007; 27(5):484-487.

(134) Winblad B, Cummings J, Andreasen N, Grossberg G, Onofrj M, Sadowsky C, Zechner S, Nagel J, Lane R. A six-month double-blind, randomized, placebo-controlled

study of a transdermal patch in Alzheimer's disease - Rivastigmine patch versus capsule. International Journal of Geriatric Psychiatry 2007; 22(5):456-467.

(135) Tariot PN, Farlow MR, Grossberg GT, Graham SM, McDonald S, Gergel I. Memantine Treatment in Patients with Moderate to Severe Alzheimer Disease Already Receiving Donepezil: A Randomized Controlled Trial. Journal of the American Medical Association 2004; 291(3):317-324.

(136) Reisberg B, Doody R, Stoffler A, Schmitt F, Ferris S, Mobius HJ. Memantine in moderate-to-severe Alzheimer's disease. New England Journal of Medicine 2003; 348(14):1333.

(137) Van Dyck CH, Tariot PN, Meyers B, Malca Resnick E. A 24-week randomized, controlled trial of memantine in patients with moderate-to-severe Alzheimer disease. Alzheimer Disease and Associated Disorders 2007; 21(2):136-143.

(138) Fuschillo C, La PS, Campana F, Pinto A, De SL. Cognitive deficits in Alzheimer's disease: Treatment with acetylcholinesterase inhibitor agents. Archives of Gerontology and Geriatrics 2001; 33:151-158.

(139) Wilkinson DG, Passmore AP, Bullock R, Hopker SW, Smith R, Potocnik FCV, Maud CM, Engelbrecht I, Hock C. A multinational, randomised, 12-week, comparative study of donepezil and rivastigmine in patients with mild to moderate Alzheimer's disease. International journal of clinical practice(Esher) 2002; 56(6):441-446.

(140) Jones RW, Soininen H, Hager K, Aarsland D, Passmore P, Murthy A, Zhang R, Bahra R. A multinational, randomised, 12-week study comparing the effects of donepezil and galantamine in patients with mild to moderate Alzheimer's disease. International Journal of Geriatric Psychiatry 2004; 19(1):58-67.

(141) Nordberg A, rreh-Shori T, Peskind E, Soininen H, Mousavi M, Eagle G, Lane R. Different cholinesterase inhibitor effects on CSF cholinesterases in Alzheimer patients. Current Alzheimer Research 2009; 6(1):4-14.

(142) Bullock R, Touchon J, Bergman H, Gambina G, He Y, Rapatz G, Nagel J, Lane R. Rivastigmine and donepezil treatment in moderate to moderately-severe

Alzheimer's disease over a 2-year period. Current Medical Research & Opinion 2005; 21(8):1317-1327.

(143) Ancoli-Israel S, Amatniek J, Ascher S, Sadik K, Ramaswamy K. Effects of galantamine versus donepezil on sleep in patients with mild to moderate Alzheimer disease and their caregivers: a double-blind, head-to-head, randomized pilot study. Alzheimer Disease & Associated Disorders 2005; 19(4):240-245.

(144) Porsteinsson AP, Grossberg GT, Mintzer J, Olin JT. Memantine treatment in patients with mild to moderate Alzheimer's disease already receiving a cholinesterase inhibitor: A randomized, double-blind, placebo-controlled trial. Current Alzheimer Research 2008; 5(1):83-89.

(145) Loveman E, Green C, Kirby J, Takeda A, Picot J, Payne E, Clegg A. The clinical and cost-effectiveness of donepezil, rivastigmine, galantamine and memantine for Alzheimer's disease. Health Technol Assess 2006; 10(1):iii-xi, 1.

(146) Evers S, Goossens M, de Vet H, van Tulder M, Ament A. Criteria list for assessment of methodological quality of economic evaluations: Consensus on Health Economic Criteria. International Journal of Technology Assessment in Health Care 2005; 21(1):240-245.

(147) Weinstein MC, O'Brien B, Hornberger J, Jackson J, Johannesson M, McCabe C, Luce BR. Principles of Good Practice for Decision Analytic Modeling in Health-Care Evaluation: Report of the ISPOR Task Force on Good Research Practice: Modeling Studies. VALUE HEALTH 2003; 6(1):9-17.

(148) Green C. Modelling disease progression in Alzheimer's disease: A review of modelling methods used for cost-effectiveness analysis. Pharmacoeconomics 2007; 25(9):735-750.

(149) Sheehan B, Phillips P, Juszczak E, Adams J, Baldwin A, Ballard C, Banerjee S, Barber B, Bentham P, Brown R, Burns A, Dening T, Findlay D, Gray R, Griffin M, Holmes C, Hughes A, Jacoby R, Johnson T, Jones R, Knapp M, Lindesay J, Mckeith I, McShane R, Macharouthu A, O'Brien J, Onions C, Passmor P, Raftery J, Ritchie C, Howard R. DOMINO-AD protocol: Donepezil and memantine in moderate to severe Alzheimer's disease - A multicentre RCT. Trials 2009; 10(57).

(150) Jonsson L, Wimo A. The Cost of Dementia in Europe A Review of the Evidence, and Methodological Considerations. Pharmacoeconomics 2009; 27(5):391-403.

(151) Oremus M. Systematic review of economic evaluations of Alzheimer's disease medications. Expert Review of Pharmacoeconomics and Outcomes Research 2008; 8(3):273-289.

(152) Weimer DL, Sager MA. Early identification and treatment of Alzheimer's disease: Social and fiscal outcomes. Alzheimers & Dementia 2009; 5(3):215-226.

(153) Getsios D, Blume S, Ishak KJ, MacLaine G. Cost-effectiveness of screening and treatment of Alzheimer's disease with donepezil in the United Kingdom. VALUE HEALTH 2009; 12(3):A191.

(154) Feldman H, Gauthier S, Hecker J, Vellas B, Hux M, Xu Y, Schwam, Shah S, Mastey V. Economic evaluation of donepezil in moderate to severe Alzheimer disease. Neurology 2004; 63(4):644-650.

(155) Fuh JL, Wang SJ, Fuh JL, Wang SJ. Cost-effectiveness analysis of donepezil for mild to moderate Alzheimer's disease in Taiwan. International Journal of Geriatric Psychiatry 2008; 23(1):73-78.

(156) Getsios D, Blume S, Ishak KJ, MacLaine G. Cost-effectiveness study in patients with mild to moderately severe Alzheimer's disease: projected benefits of donepezil in the United Kingdom. VALUE HEALTH 2009; 12(3):A191.

(157) Getsios D, Blume S, Ishak KJ, MacLaine G. Cost-effectiveness of donepezil in the treatment of mild to moderate Alzheimer's disease: a UK evaluation using Discrete-Event Simulation. Pharmacoeconomics 2010; 28(5):411-427.

(158) Lopez-Bastida J, Hart W, Garcia-Perez L, Linertova R. Cost-effectiveness of donepezil in the treatment of mild or moderate Alzheimer's disease. Journal of Alzheimer's Disease 2009; 16(2):399-407.

(159) Iu S, Hill J, Fillit H. Impact of donepezil use in routine clinical practice on health care costs in patients with Alzheimer's disease and related dementias enrolled in a

large medicare managed care plan: A case-control study. American Journal Geriatric Pharmacotherapy 2005; 3(2):92-102.

(160) Mesterton J, By A, Sandelin R, Jonsson L. Cost-effectiveness of donepezil in Alzheimer's disease in Sweden. VALUE HEALTH 2009; 12(7):A369.

(161) Pattanaprateep O, Phongchareonsuk P, Chaikledkaew U. The cost-effectiveness of donepezil and rivastigmine in the treatment of Alzheimer's disease in Thailand private hospital. VALUE HEALTH 2005; 8(3):314-315.

(162) Teipel SJ, Ewers M, Reisig V, Schweikert B, Hampel H, Happich M. Long-term cost-effectiveness of donepezil for the treatment of Alzheimer's disease. European Archives of Psychiatry and Clinical Neuroscience 2007; 257(6):330-336.

(163) Feldman H, Gauthier S, Hecker J, Vellas B, Hux M, Xu Y, Schwam E, Leader M, Shah SN. Pharmacoeconomic benefits of donepezil treatment in severe Alzheimer's disease. Journal of the American Geriatrics Society 2004; 52(4):388.

(164) Shah SN, Feldman H, Gauthier S, Hecker J, Vellas B, Hux M, Xu Y, Schwam E, Leaderer M. Pharmacoeconomic benefits of donepezil treatment in severe Alzheimer's disease. Neurobiology of Aging 2004; 25(Suppl. 2):S208.

(165) Feldman H, Hux M, Schwam EM. Economic evaluation of donepezil in moderate to severe Alzheimer disease - Reply. Neurology 2005; 64(7):1320.

(166) Faltraco F, Happich M, Reisig V, Teipel SJ, Moeller HJ, Hampel H. Treatment of Alzheimer's disease with cholinesterase inhibitors in Germany: A cost-effectiveness study. Neurobiology of Aging 2004; 25(Suppl. 2):S336.

(167) Gauthier S, Feldman H, Hecker J, Vellas B, Ames D, Subbiah P, Whalen E, Emir B. Efficacy of donepezil on behavioral symptoms in patients with moderate to severe Alzheimer's disease. Int Psychogeriatr 2002; 14(4):389-404.

(168) Brennan A, Nagy B, Brandtmuller A, Thomas SK, Sullivan SD, Akehurst R. The cost–utility of exelon patch in the management of patients with moderate Alzheimer's disease in the United Kingdom. VALUE HEALTH 2007; 10(6):A384.

(169) Suh GH, Jung HY, Lee CU, Choi S. Economic and clinical benefits of galantamine in the treatment of mild to moderate Alzheimer's disease in a Korean population: a 52-week prospective study. Journal of Korean Medical Science 2008; 23(1):10-17.

(170) Suh GH. Modeling the cost-effectiveness of galantamine for mild to moderately severe Alzheimer's disease in Korea. VALUE HEALTH 2009; 12(SUPPL. 3):S49-S54.

(171) Antonanzas F, Rive B, Badenas JM, Gomez-Lus S, Guilhaume C. Cost-effectiveness of memantine in community-based Alzheimer's disease patients: An adaptation in Spain. European Journal of Health Economics 2006; 7(2):137-144.

(172) Gagnon M, Rive B, Hux M, Guilhaume C. Cost-effectiveness of memantine compared with standard care in moderate-to-severe Alzheimer disease in Canada. Canadian Journal of Psychiatry 2007; 52(8):519-526.

(173) Guilhaume C, Rive B, Francois C, Livingston G, Katona C. External validation of the probabilistic Markov model estimating the cost effectiveness of memantine versus standard care in Alzheimer disease from a UK perspective. VALUE HEALTH 2005; 8(6):A128.

(174) Jonsson L. Cost-effectiveness of memantine for moderate to severe Alzheimer's disease in Sweden. American Journal Geriatric Pharmacotherapy 2005; 3(2):77-86.

(175) Toumi M, Lamure M, Grishchenko M, Cochran J, Rive B. Cost-effectiveness of memantine in the treatment of moderate to severe Alzheimer's disease in Norway. Value Health 2009; 12(7):A370.

(176) Weycker D, Taneja C, Edelsberg J, Erder MH, Schmitt FA, Setyawan J, Oster G. Cost-effectiveness of memantine in moderate-to-severe Alzheimer's disease patients receiving donepezil. Current Medical Research and Opinion 2007; 23(5):1187-1197.

(177) Iglehart D. Simulating Stable Stochasitic Systems, V: Comparison of Ratio Estimators. Naval Res Logist Quart 1975; 22:553-565.

(178) Kavanagh S, Knapp M. Costs and cognitive disability: modelling the underlying associations. British Journal of Psychiatry 2002; 180:120-125.

(179) Lowin A, Knapp M, McCrone P. Alzheimer's disease in the UK: comparative evidence on cost of illness and volume of health services research funding. International Journal of Geriatric Psychiatry 2001; 16:1143-1148.

(180) Souêtre E, Thwaites R, Yeardley H. Economic impact of Alzheimer's disease in the United Kingdom: Cost of care and disease deverity for non-institutionaalised patients with Alzheimer's disease. British Journal of Psychiatry 1999; 174:51-55.

(181) Wolstenholme J, Fenn P, Gray A, Keene J, Jacoby R, Hope T. Estimating the relationship between disease progression and cost of care in dementia. British Journal of Psychiatry 2002; 181:36-42.

(182) Livingston G, Katona C, Roch B, Guilhaume C, Rive B. A dependency model for patients with Alzheimer's disease: its validation and relationship to the costs of care - the LASER-AD study. Current Medical Research and Opinion 2004; 20(7):1007-1016.

(183) Jonsson L, Wimo A. The Cost of Dementia in Europe A Review of the Evidence, and Methodological Considerations. Pharmacoeconomics 2009; 27(5):391-403.

(184) Quentin W, Riedel-Heller S, Luppa M, Rudolph A, König H-H. Cost-of illness studies of dementia: a systematic review focusing on stage dependency of costs. Acta Psychiatrica Scandinavica 2009.

(185) Hatoum HT, Thomas SK, Lin SJ, Lane R, Bullock R. Predicting time to nursing home placement based on activities of daily living scores - A modelling analysis using data on Alzheimer's disease patients receiving rivastigmine or donepezil. Journal of Medical Economics 2009; 12(2):98-103.

(186) Miller EA, Schneider LS, Rosenheck RA. Assessing the relationship between health utilities, quality of life, and health services use in Alzheimer's disease. International Journal of Geriatric Psychiatry 2009; 24(1):96-105.

(187) Zhu CW, Leibman C, McLaughlin T, Scarmeas N, Albert M, Brandt J, Blacker D, Sano M, Stern Y. The effects of patient function and dependence on costs of care in Alzheimer's disease. Journal of the American Geriatrics Society 2008; 56(8):1497-1503.

(188) Getsios D, Migliaccio-Walle K, Caro JJ. NICE cost-effectiveness appraisal of cholinesterase inhibitors - Was the right question posed? Were the best tools used? Pharmacoeconomics 2007; 25(12):997-1006.

(189) Cohen JT, Neumann PJ. Decision analytic models for Alzheimer's disease: State of the art and future directions. Alzheimers & Dementia 2008; 4(3):212-222.

(190) Green C. Modelling disease progression in Alzheimer's disease - A review of modelling methods used for cost-effectiveness analysis. Pharmacoeconomics 2007; 25(9):735-750.

(191) Stern Y, Tang MX, Albert MS, Brandt J, Jacobs DM, Bell K, Marder K, Sano M, Devanand D, Albert SM, Bylsma F, Tsai WY. Predicting time to nursing home care and death in individuals with Alzheimer disease. JAMA 1997; 277(10):806-812.

(192) Stern Y, Albert SM, Sano M, Richards M, Miller L, Folstein M, Albert M, Bylsma FW, Lafleche G. Assessing patient dependence in Alzheimer's disease. Journal of Gerontology 1994; 49(5):216-222.

(193) Hoyle M, Anderson R. Whose costs and benefits? Why economic evaluations should simulate both prevalent and all future incident patient cohorts. Medical Decision Making 2010; *In press*.

(194) Livingston G, Katona C, Francois C, Guilhaume C, Cochran J, Sapin C. Characteristics and health status change over 6 months in people with moderately severe to severe Alzheimer's disease in the UK. International Psychogeriatrics 2006; 18(3):527-538.

(195) National Institute for Health and Clinical Excellence. Guide to the Methods of Technology Appraisal. 2008. London, National Institute for Clinical Excellence.

(196) Galasko D, Schmitt F, Thomas R, Jin S, Bennett D, Ferris S. Detailed assessment of activities of daily living in moderate to severe Alzheimer's disease. Journal of the International Neuropsychological Society 2005; 11(4):446-453.

(197) Briggs A, Claxton K, Sculpher M. Decision modelling for health economic evaluation. Oxford: Oxford University Press; 2008.

(198) Doraiswamy PM, Bieber F, Kaiser L, Krishnan KR, Reuning-Scherer J, Gulanski B. The Alzheimer's Disease Assessment Scale: patterns and predictors of baseline cognitive performance in multicenter Alzheimer's disease trials. Neurology 1997; 48(6):1511-1517.

(199) Miller EA, Schneider LS, Zbrozek A, Rosenheck RA. Sociodemographic and Clinical Correlates of Utility Scores in Alzheimer's Disease. Value Health 2008; 11(7):1120-1130.

(200) Naglie G, Tomlinson G, Tansey C, Irvine J, Ritvo P, Black SE, Freedman M, Silberfeld M, Krahn M. Utility-based quality of life measures in Alzheimer's disease. Quality of Life Research 2006; 15(4):631-643.

(201) Jonsson L, Andreasen N, Kilander L, Soininen H, Waldemar G, Nygaard H, Winblad B, Jonhagen ME, Hallikainen M, Wimo Al. Patient- and proxy-reported utility in Alzheimer disease using the EuroQoL. Alzheimer Disease & Associated Disorders 2006; 20(1):49-55.

(202) Ekman M, Berg J, Wimo A, Jonsson L, McBurney C. Health utilities in mild cognitive impairment and dementia: a population study in Sweden. International Journal of Geriatric Psychiatry 2007; 22:649-655.

(203) Wlodarczyk JH, Brodaty H, Hawthorne G, Wlodarczyk JH, Brodaty H, Hawthorne G. The relationship between quality of life, Mini-Mental State Examination, and the Instrumental Activities of Daily Living in patients with Alzheimer's disease. Archives of Gerontology & Geriatrics 2004; 39(1):25-33.

(204) Andersen CK, Wittrup-Jensen KU, Lolk A, Andersen K, Kragh S. Ability to perform activities of daily living is the main factor affecting quality of life in patients with dementia. Health & Quality of Life Outcomes 2004; 2:52.

(205) Karlawish JH, Zbrozek A, Kinosian B, Gregory A, Ferguson A, Low DV, Glick HA. Caregivers' assessments of preference-based quality of life in Alzheimer's disease. Alzheimer's and Dementia 2008; 4(3):203-211.

(206) Karlawish JH, Zbrozek A, Kinosian B, Gregory A, Ferguson A, Glick HA,. Preference-based quality of life in patients with Alzheimer's disease. Alzheimer's & Dementia 2008; 4(3):193-202.

(207) Kerner DN, Patterson TL, Grant I, Kaplan RM. Validity of the Quality of Well-Being Scale for Patients with Alzheimer's Disease. J Aging Health 1998; 10(1):44-61.

(208) Karlawish JH, Casarett D, Klocinski J, Clark CM. The relationship between caregivers' global ratings of Alzheimer's disease patients' quality of life, disease severity, and the caregiving experience. Journal of the American Geriatrics Society 2001; 49:1066-1070.

(209) Sands LP, Ferreira P, Stewart AL, Brod M, Yaffe K. What explains differences between dementia patients' and their caregivers' ratings of patients' quality of life? American Journal of Geriatric Psychiatry 2004; 12(3):272-280.

(210) Vogel A, Mortensen EL, Hasselbalch SG, Andersen BB, Waldemar G. Patient versus informant reported quality of life in the earliest phases of Alzheimer's disease. International Journal of Geriatric Psychiatry 2006; 21(12):1132-1138.

(211) Huang HL, Chang MY, Tang JSH, Chiu YC, Weng LC. Determinants of the discrepancy in patient- and caregiver-rated quality of life for persons with dementia. Journal of Clinical Nursing 2008; 18:3107-3117.

(212) Wlodarczyk JH, Brodaty H, Hawthorne G. The relationship between quality of life, Mini-Mental State Examination, and the Instrumental Activities of Daily Living in patients with Alzheimer's disease. Archives of Gerontology and Geriatrics 2004; 39(1):25-33.

(213) Neumann PJ, Kuntz KM, Leon J, Araki SS, Hermann RC, Hsu MA, Weinstien MC. Health utilities in Alzheimer's disease: a cross-sectional study of patients and caregivers. Med Care 1999; 37(1):27-32.

(214) Perneczky R, Wagenpfeil S, Komossa K, Grimmer T, Diehl J, Kurz A. Mapping scores onto stages: mini-mental state examination and clinical dementia rating. American Journal of Geriatric Psychiatry 2006; 14(2):139-144.

(215) Curtis L. Unit Costs of Health & Social Care 2009. 2009. Canterbury, PSSRU, University of Kent.

(216) Curtis L, Netten A. Unit Costs of Health & Social Care 2004. 2004. Canterbury, PSSRU, University of Kent.

(217) Hoyle M. Future drug prices and cost-effectiveness analyses. Pharmacoeconomics 2008; 26(7):589-602.