The effectiveness and cost-effectiveness of donepezil, galantamine, rivastigmine and memantine for the treatment of Alzheimer's disease (review of Technology Appraisal No. 111): a systematic review and economic model

M Bond,[1]* G Rogers,[1] J Peters,[1] R Anderson,[1] M Hoyle,[1] A Miners,[2] T Moxham,[1] S Davis,[3] P Thokala,[3] A Wailoo,[3] M Jeffreys[4] and C Hyde[1]

[1]Peninsula Technology Assessment Group (PenTAG), University of Exeter, Exeter, UK
[2]London School of Hygiene and Tropical Medicine, London, UK
[3]The School of Health and Related Research (ScHARR), University of Sheffield, Sheffield, UK
[4]Royal Devon and Exeter Foundation Trust Hospital, Exeter, UK

*Corresponding author

Executive summary

**Health Technology Assessment
NIHR HTA programme
www.hta.ac.uk**

Executive summary

Background

Alzheimer's disease (AD) is the most commonly occurring form of dementia, accounting for approximately 62% of instances of dementia. AD is predominantly a disease of later life, with 5% of the UK population over 65 years affected. In England and Wales, among people aged 65–69 years, the incidence is estimated to be 7.4 [95% confidence interval (CI) 3.6 to 16.1] per 1000 person-years, rising to 84.9 (95% CI 63.0 to 107.8) per 1000 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 AD (see above), then there are approximately 111,600 new cases in England and Wales per year.

Methods

Interventions

This technology assessment report (TAR) considered four interventions. Three have UK marketing authorisations for the treatment of adults with mild-to-moderately severe AD [Mini Mental State Examination (MMSE) score 26–10]. These are donepezil (Aricept®, manufactured by Eisai Ltd), rivastigmine (Exelon®, manufactured by Novartis) and galantamine (Reminyl®, manufactured by Shire Pharma). They are acetylcholinesterase inhibitors (AChEIs). The fourth drug, memantine hydrochloride (Ebixa®), manufactured by Lundbeck, has a UK marketing authorisation for the treatment of moderate-to-severe AD (MMSE ≤ 20). It is a voltage-dependent, moderate-affinity, uncompetitive N-methyl-D-aspartate receptor antagonist.

Comparators

The comparators are mild AD (MMSE 21–26) – donepezil, galantamine and rivastigmine; moderate AD (MMSE 10–20) – donepezil, galantamine, rivastigmine and memantine; and severe AD (MMSE < 10) – memantine. All of the above were also compared with best supportive care (BSC) (i.e. without treatment with any AChEIs or memantine).

Population

The population is adults with AD.

Outcome measures

The outcomes include:

- 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 (HRQoL) of patients and carers
- adverse effects of treatment
- cost-effectiveness and costs.

Study design

Systematic review of clinical effectiveness: only systematic reviews of randomised controlled trials (RCTs) and RCTs were considered.

Systematic review of economic evaluations: designs were identical to those for the systematic review of clinical effectiveness, except non-randomised studies, full cost-effectiveness analyses, cost–utility analyses, cost–benefit analyses, cost–consequence analyses and stand-alone cost analyses based in the UK NHS were included.

Clinical effectiveness systematic review

Data sources

Electronic databases were searched in November 2009 and updated in March 2010; this updated search revealed no new includable studies. The search strategy can be found in *Appendix 2*. The databases searched included The Cochrane Library (2009 Issue 4, Cochrane Database of Systematic Reviews and Cochrane Central Register of Controlled Trials), MEDLINE, MEDLINE In-Process & Other Non-Indexed Citations, EMBASE, PsycINFO, EconLit, ISI Web of Science Databases – Science Citation Index, Conference Proceedings Citation Index, and BIOSIS; the Centre for Reviews and Dissemination (CRD) databases – NHS Economic Evaluation Database, Health Technology Assessment and Database of Abstracts of Reviews of Effects databases. Where possible, a controlled trials and human filter was added. As this was an update of a previous review, the searches were run from 2004 to present. The meta-register of controlled trials and 'clinicaltrials.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. As a result of resource limitations the search was restricted to English-language papers only.

Study selection

Relevant studies were identified in two stages. Titles and abstracts were examined independently by two researchers and screened for possible inclusion. Disagreements were resolved by discussion. Full texts of the identified studies were obtained. Two researchers examined these independently for inclusion or exclusion, and disagreements were resolved by discussion. A third reviewer was available if necessary.

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: pair-wise meta-analysis, pooling of multiple outcome measures, metaregression and mixed-treatment comparisons – indirect comparison.

Review of past economic evaluations

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 stand-alone cost analyses based in the UK NHS. Narrative synthesis, supported by the data extraction tables, was used to summarise the evidence base.

Peninsula Technology Assessment Group cost–utility model

A decision model based broadly on the structure of the three-state Markov model described in the previous TAR was developed, based upon time to institutionalisation and parameterised with updated estimates of effectiveness, costs and utilities. For the three cholinesterase inhibitors the base-case analysis modelled a cohort of people with mild-to-moderate AD. For memantine, the base-case analysis concerned people with moderate-to-severe AD. Exploratory sensitivity analyses looked at people with only mild, moderate or severe AD.

Disease progression based on age, MMSE and uniform Activities of Daily Living (ADL) Scale was modelled using individual patient data (IPD) from the UK-based study by Wolstenholme and colleagues (2002) (patient data from 1988 to 1999 in Oxfordshire). The study data supplied by Wolstenholme and colleagues also provided estimates of the NHS and Personal Social Services (PSS) costs associated with AD. Data from the LASER-AD longitudinal cohort study were also used to justify and/or corroborate a number of assumptions within the model.

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.

Clinical effectiveness results

Number and quality of effectiveness studies

From 1843 titles and abstracts screened, four systematic reviews and 17 RCTs were found which matched our inclusion criteria that had been published since 2004. There were 12 pair-wise 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.

Summary of benefits and risks

Donepezil: our systematic review found five small poor-quality studies that have added to the evidence base. All studies measured cognitive outcomes. A dose-related beneficial effect was found at 10 mg/day.

Galantamine: we found an additional three variable-quality RCTs of galantamine versus placebo to add to the evidence base of six studies included in 2004. 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 with 2004 evidence, significant gains for people taking galantamine were found for cognitive, functional and global outcomes.

Rivastigmine: our update review found three new studies; one of these was of reasonable size and quality. Positive benefits from rivastigmine were found on cognitive, functional and global outcomes, but, as in 2004, 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.

Memantine: we found a new, poorer-quality study which failed to show any benefit from memantine on any outcome measure. When the data were pooled with 2004 evidence, 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. 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 found statistically significant differences only on functional and global outcomes, both favouring rivastigmine. One new study and one previous study compared donepezil with galantamine; neither was of good quality. The new study only looked at global outcomes and found no difference between the treatments. Finally, 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 AChEI against AChEI 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 an underlying pharmacological interaction between galantamine and memantine – which neutralises their effects – in the new trial, which used all three AChEIs, whereas the existing trial only combined memantine with donepezil. The other difference between these studies is the lack of ITT in the former one, which may have led to more favourable results for combination therapy.

Cost-effectiveness results

Published economic evaluations

The systematic review of economic evaluations identified 23 included studies, over one-third of which were published only as abstracts and could not be considered in depth. Of the remainder, most studies 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.

Peninsula Technology Assessment Group modelling results

Notwithstanding the uncertainty about our findings, in contrast with the previous TAR, we found in the base case that the AChEIs are probably cost saving at a willingness-to-pay (WTP) threshold of £30,000 per QALY for people with mild-to-moderate AD. For this class of drugs, there is a > 99% probability that the AChEIs are more cost-effective than BSC. These analyses assume that the AChEIs have no effect on survival. If a survival effect is assumed, the AChEIs no longer dominate BSC, and incremental cost-effectiveness ratios (ICERs) for the AChEIs are approximately £37,000. However, as we have not been able to find any relevant studies that measure survival, these ICERs are purely speculative.

For the AChEIs, in people with mild-to-moderate AD, the probabilistic sensitivity analyses suggested that donepezil is the most cost-effective, but with a probability of only 28% of being the most cost-effective option at a WTP of £30,000 per QALY (27% at a WTP of £20,000 per QALY). In the deterministic results, donepezil dominates the other drugs and BSC, which, along with rivastigmine patches, are associated with greater costs and fewer QALYs. Thus, although galantamine has a slightly cheaper total cost than donepezil (£69,592 vs £69,624), the slightly greater QALY gains from donepezil (1.616 vs 1.617) are enough for donepezil to dominate galantamine.

The probability that memantine is cost-effective in a moderate-to-severe cohort compared with BSC [see *Chapter 6, Moderate to severe Alzheimer's disease: memantine (Decision problem 2a)*] at a WTP of £30,000 per QALY is 38% (and 28% at a WTP of £20,000 per QALY). The deterministic ICER for memantine is £32,100 per/QALY and the probabilistic ICER is £36,700 per/QALY. Sensitivity analyses, assuming that memantine gave an additional 1.7 months of life, changed the ICER to £65,619 per QALY, owing to the modest utility gains and greater additional cost.

Discussion

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 6 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, institutionalisation, the impact on carer's time and the prescription of antipsychotics.
■ None of the trials conducted subgroup 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 a 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 AD (AD Assessment Scale – Cognitive Subscale, MMSE). Therefore, the effects of treatment may have been underestimated in some cases.
■ The searches were limited to the English language owing to resource limitations, which may have led us to exclude important studies.

Strengths and limitations of the economic modelling by the Peninsula Technology Assessment Group

We have made a number of improvements to the previous SHTAC-AHEAD (Southampton Health Technology Assessment Centre –Assessment of Health Economics in Alzheimer's disease model) and attempted to address some of the specific criticisms of the previous model. However, some limitations remain:

■ The underlying disease model captures just the dimensions of cognitive status and functional status/ADL. Behavioural and psychological symptoms are not incorporated into the model, and, therefore, any treatment effects and QoL impacts related to these symptoms will not be captured.
■ The expression of treatment effectiveness, although based on a multivariate formula based on patient age, ADL and cognitive status, is mainly based on predicting delays in time to institutionalisation. Although 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 QoL is uncertain.
■ Although the model now incorporates more graduated declines in patient utility, and more graduated increases in NHS and PSS costs prior to institutionalisation, assuming that all of these time-related cost and utility changes will be delayed by the same amount of time that institutionalisation is delayed is a key assumption in the model.
■ The main database of IPD from the UK that the time-to-institutionalisation model and key cost parameters are largely based upon is relatively old (1988–99), small ($n = 92$ with AD) and from one county of the UK (Oxfordshire). Its generalisability to England and Wales in 2010, therefore, has to be considered.
■ 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 that estimating the benefits of drug treatments for those

already in residential care is problematic. This is a more considerable weakness for evaluating the cost-effectiveness of memantine.

■ The generalisability of the UK IPD should be questioned because (1) the data are from just 92 individuals; (2) they were collected from Oxfordshire only; and (3) these data are now rather out of date, as they were collected between 1988–9 and 1999. This may impact on the generalisability of the model.

■ The incorporation of the full treatment effect at 6 months is artificial. It is more likely that improvements due to treatment are gradual. It is also assumed in the Peninsula Technology Assessment Group (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 *Appendix 7*.

■ 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).

Conclusions

The additional clinical effectiveness evidence identified in this update systematic review continues to suggest clinical benefit from the AChEIs in alleviating AD symptoms, although there is considerable debate about the magnitude of the effect. There is also some evidence that AChEIs have an impact on controlling disease progression. Although there is also new evidence on the effectiveness of memantine, it remains less supportive of this drug's use than the evidence for AChEIs.

The conclusions concerning cost-effectiveness are quite different from the previous assessment. This is because both the changes in effectiveness and costs between drug use and non-drug use underlying the ICERs are very small. This leads to highly uncertain results, which are very sensitive to change.

Implications for service provision

These are not clear and will ultimately rest on the interpretation of the new evidence.

Suggested research priorities

■ Good-quality longer-term RCTs (following CONSORT; consolidated standards of reporting trials) to include mortality, time to institutionalisation and HRQoL as outcomes and sufficiently powered for subgroup analysis by disease severity, response to treatment, behavioural disturbance and comorbidities.

■ 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 measures validated for people with dementia, for example DEMQOL. Work is needed to derive utility values from such validated measures.

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

■   Research into more valid ways of accounting for missing data than LOCF and OC, particularly in degenerative diseases such as AD.

Funding

Publication

NIHR Health Technology Assessment programme

The Health Technology Assessment (HTA) programme, part of the National Institute for Health Research (NIHR), was set up in 1993. It produces high-quality research information on the effectiveness, costs and broader impact of health technologies for those who use, manage and provide care in the NHS. 'Health technologies' are broadly defined as all interventions used to promote health, prevent and treat disease, and improve rehabilitation and long-term care.

The research findings from the HTA programme directly influence decision-making bodies such as the National Institute for Health and Clinical Excellence (NICE) and the National Screening Committee (NSC). HTA findings also help to improve the quality of clinical practice in the NHS indirectly in that they form a key component of the 'National Knowledge Service'.

The HTA programme is needs led in that it fills gaps in the evidence needed by the NHS. There are three routes to the start of projects.

First is the commissioned route. Suggestions for research are actively sought from people working in the NHS, from the public and consumer groups and from professional bodies such as royal colleges and NHS trusts. These suggestions are carefully prioritised by panels of independent experts (including NHS service users). The HTA programme then commissions the research by competitive tender.

Second, the HTA programme provides grants for clinical trials for researchers who identify research questions. These are assessed for importance to patients and the NHS, and scientific rigour.

Third, through its Technology Assessment Report (TAR) call-off contract, the HTA programme commissions bespoke reports, principally for NICE, but also for other policy-makers. TARs bring together evidence on the value of specific technologies.

Some HTA research projects, including TARs, may take only months, others need several years. They can cost from as little as £40,000 to over £1 million, and may involve synthesising existing evidence, undertaking a trial, or other research collecting new data to answer a research problem.

The final reports from HTA projects are peer reviewed by a number of independent expert referees before publication in the widely read journal series *Health Technology Assessment*.

Criteria for inclusion in the HTA journal series
Reports are published in the HTA journal series if (1) they have resulted from work for the HTA programme, and (2) they are of a sufficiently high scientific quality as assessed by the referees and editors.

Reviews in *Health Technology Assessment* are termed 'systematic' when the account of the search, appraisal and synthesis methods (to minimise biases and random errors) would, in theory, permit the replication of the review by others.