MEMANTINE FOR THE TREATMENT OF MODERATE TO SEVERE ALZHEIMER'S DISEASE

Lundbeck Response to the Technology Assessment Report

4th August 2010

Abbreviations

AChEI	Acetylcholinesterase inhibitor
AD	Alzheimer's disease
ADAS-Cog	Alzheimer's disease assessment scale – cognitive subscale
ADCS-ADL	Alzheimer disease cooperative study – activities of daily living
AHEAD	Assessment of health economics in Alzheimer's disease
AE	Adverse event
APS	Agitation/aggression and/or psychotic symptoms
CHMP	Committee for Medicinal Products for Human Use
EMA	European medicines agency
FDA	Food and drug administration
GCP	Good clinical practice
GDS	Global deterioration scale
ICERs	Incremental cost-effectiveness ratios
ICH	International conference on harmonisation
IPD	Individual patient data
ITT	Intention to treat
LOCF	Last observation carried forward
MMSE	Mini-mental state examination
MTA	Multiple technology appraisal
MTC	Mixed treatment comparison
NICE	National institute for health and clinical excellence
NPI	Neuropsychiatric inventory
OC	Observed cases
PenTAG	Peninsula technology assessment group
PSSRU	Personal social services research unit
QALY	Quality-adjusted life years
RCT	Randomised controlled trial
SHTAC	Southampton technology assessment centre

SIB	Severe impairment battery
SMD	Standard mean difference
TAR	Technology assessment review
UK	United Kingdom
WMD	Weighted mean difference

Table of Contents

1 Executive Summary

Lundbeck has two major reservations about the PenTAG evaluation and the TAR conclusions:

- First, the PenTAG model used to derive cost-effectiveness estimates in this report is flawed in several major respects. Not only is it based on a number of incorrect assumptions and a very limited dataset, it also lacks both face and technical validity. The economic evaluation undertaken by PenTAG was poorly executed; the model underestimates the treatment benefits of memantine and increases the cost-effectiveness ratio estimate.

- Second, apart from the model used, the approach taken by PenTAG in their review of the clinical evidence for memantine meant that vital evidence was excluded and memantine was not evaluated across its full licensed indication. Importantly, this was not clearly stated in the TAR. Only evidence in moderately-severe to severe AD was included. Evidence for memantine in patients with moderate AD was omitted from the review, despite the availability of this data in the public domain. Furthermore, unpublished evidence submitted by Lundbeck in a sub-group population with behavioural disturbances was not included in the review.

The cost-effectiveness estimate of memantine derived from the current PenTAG model is misleading and should not be relied upon to make recommendations on the use of memantine in England and Wales. The use of an inappropriate economic model has serious consequences and could lead to many thousands of patients being denied cost-effective interventions. Moreover, because of the reluctance of PenTAG to consider the six memantine RCTs submitted for review and the flawed development of the model, memantine could never demonstrate clinical effectiveness and cost-effectiveness in either the licensed population of moderate to severe AD patients, or in the behavioural sub-group of patients with APS. Therefore, we urge the Appraisal Committee to consider using the health economic model developed by Lundbeck, which has been proven to be robust and well calibrated for moderate and severe AD.

This Executive Summary summarises our major concerns with the PenTAG evaluation by considering the validity of the new economic model developed by the PenTAG group, PenTAG's critique on the Lundbeck model, the critical limitations of the clinical effectiveness evidence review, and finally the overall approach of PenTAG to the evaluation within this MTA. Additional information to support this summary is provided in the main body of the response and corresponding appendices.

The PenTAG Model

1. The PenTAG model is flawed in several major respects. As a result, the cost-effectiveness estimates reported are invalid.

- The PenTAG model excludes the behavioural symptom domain of AD – this is despite explicit acknowledgement in the background text that behavioural symptoms have a considerable impact on disease progression, and specifically on time to institutionalisation, and should therefore be included in any robust model of AD.

- The assumptions used in the predictive equation developed for the model are methodologically flawed and collectively these not only underestimate the clinical effectiveness of memantine and the AChEIs, but also exaggerate their cost-effectiveness estimates:
  1. The assumption that hazard rates would not accelerate over time has no clinical validity.
  2. The model relies on the uncertain relationship between institutionalisation and cognition and global function. In the PenTAG predictive equation age is the only significant predictor of time to institutionalisation; cognition and global function were forced into the model, despite being found to be insignificant predictors. Conversely, the important and significant predictors identified in the original study analysis (e.g. gender, education, presence of behavioural symptoms, living arrangements) were ignored. This is not in accordance with the published literature. The small and non-significant impact of cognition and function on institutionalisation in the PenTAG equation reflects the inability of the model to capture changes on either of these variables. Because patient's age was the only influential predictor of institutionalisation in the PenTAG model, the model cannot produce meaningful estimates of treatment effects since clearly no treatment can have an impact on age. This makes the high ICERs for memantine a foregone conclusion.
  3. The model assumes that there is no onset of a treatment effect until six months. This assumption has no clinical basis and contradicts the RCT evidence. This leads to an underestimation of the true benefit of the therapies within the model.
  4. The data collected by Wolstenholme and colleagues, on which the model is developed, has several limitations: it is outdated, comprised only 92 untreated patients and it did not use a standard functional disability scale. The limitations of converting the ADCS-ADL scales into the Barthel index resulted in an underestimation of the benefit of AD medications on the functional domain.
  5. Memantine data was inappropriately handled when assessing the effect on the Barthel index. This resulted in an additional underestimation of memantine benefit on functioning.
  6. Further unjustified assumptions were made that effectively disable the model's ability to show QALY gains due to prolonged time to institutionalisation in patients with moderate and severe AD:
     - 40% of moderate to severe patients were assumed to be institutionalised from the model start.
     - The utility values for institutionalised patients were assumed to be equal to the values for patients with severe AD who are not institutionalised.

- These assumptions further undermine the validity of the PenTAG cost-effectiveness assessment of memantine.

- Reporting on the model is of a poor quality. The statistical procedure used to build the equation was poorly discussed and reported. Neither the results of the univariate analyses, the testing of relationships between all predictors and the outcome variable (institutionalisation), nor the selection process of the model were described. No information regarding uncertainty of the equation coefficients (p-values, standard errors or confidence intervals) was reported. The true predictors of time to institutionalisation based on the utilised database are unclear. Additionally no data are presented to assess how these true predictors could potentially impact the final model when included in the equation along with the insignificant predictors (MMSE, Barthel index).

- Given the multiple issues with face validity described above, it is somewhat moot to consider the technical validity of the model. However multiple technical issues were identified:
  - PenTAG did not implement the simple model concept as specified. The model does not represent a Markov structure as there were no transitions between the four states ("pre-FTC/inst", "FTC/inst", "alive", and "dead").
  - PenTAG implemented a modelling structure that purports to allow for diminishing utility and rising costs over time thereby addressing limitations of the previous model. However, an inspection of the model reveals that the calculations do not implement these changes in cost and utility and that the calculations used in fact generate values that vary considerably from the source data. It should also be noted that in the model utilities vary with age although no such relationship was postulated.
  - The model does not behave correctly as exemplified by the fact that small changes in mortality lead to negative numbers in the 'institution' case leading to the total costs being lower than the pre-institutionalisation costs.

- The PenTAG model has not been validated in any way, nor has it been published in a peer-reviewed journal. PenTAG rejected several other published modelling approaches in favour of developing their own model.

2. It is of particular note that the PenTAG model used in this review is inappropriate for an evaluation of the cost-effectiveness of memantine in moderate and severe AD.

- The model excluded the behavioural domain, which is particularly relevant when predicting the time to institutionalisation in moderate and severe AD.
- The PenTAG model was built on data derived from an untreated cohort of AD patients. While this represents a suitable approach to evaluate the cost-effectiveness of AChEIs, the cost-effectiveness of memantine should be evaluated in the target population for memantine. Routine clinical practice in the UK is such that the majority of moderate and severe AD patients will have received AChEIs at some point in the course of the disease.

The Lundbeck Model

3. The model submitted by Lundbeck adopted a robust and well calibrated approach to evaluate the cost-effectiveness of memantine for the licensed patient population.

- The Lundbeck model adopted the framework developed for the previous technology appraisal in 2004, but with improvements to address the key limitations of the SHTAC-AHEAD model.
- The predictive equation used in the cost-effectiveness analysis submitted by Lundbeck included all three core domains of AD (cognition, functioning and behaviour) identified by PenTAG as being necessary for any comprehensive economic evaluation in AD.
- The Lundbeck model was built using data from the LASER-AD database, which represents a cohort of AD patients from the UK that best reflects routine clinical practice. These patients are also exactly aligned with the licensed population for memantine.
- The Lundbeck methodology for the development of this new predictive equation and its application in evaluating the cost-effectiveness of memantine in the UK clinical setting has been published in peer-reviewed journals.[1,2]
- PenTAG made criticisms of the Lundbeck model, however none of these were major, nor necessitated extensive additional analyses, structural changes or resulted in a change to the model conclusions.

Lundbeck believe that substantial work must be undertaken before the current PenTAG model is fit-for-purpose to ensure estimates of the cost-effectiveness of AD treatments are reasonably accurate, meaningful and realistic enough to inform decisions about the appropriate use of NHS resources.

As there were no major criticisms by PenTAG of the model submitted by Lundbeck, and none that would result in a change to the model conclusions, it is proposed that on this occasion the two distinct classes of drugs in AD should be assessed using different models, and that memantine is evaluated using the model developed by Lundbeck.

[1] Rive B, Le Reun C, Grishchenko M, et al., 2010a. Predicting time to full-time care in AD: a new model. JME, 13(2): 362-70.

[2] Rive B, Grishchenko M, Guilhaume-Goulant C, et al., 2010b. Cost-effectiveness of Memantine in Alzheimer's Disease in the UK. JME, 13(2): 371-80.

The Clinical Effectiveness of Memantine

4. PenTAG's conclusions on the clinical effectiveness of memantine were limited to a review of two studies in patients with moderately-severe to severe AD. More recent studies including patients with moderate AD were inappropriately excluded from the review, as was an EMA-approved published meta-analysis of six RCTs. PenTAG therefore failed to address the decision problem outlined in the final scope and failed to review memantine in its full licensed indication.

- The remit assigned to PenTAG was to review the effectiveness of the AD therapies within their licensed indications. For memantine this includes all patients with moderate to severe AD (MMSE <20), so comprises those who have previously received AChEIs, those naive to AD medication and those who may currently receive a stable dose of an AChEI.
- Lundbeck provided evidence for the clinical efficacy of memantine based on a published meta-analysis of six large multicentre RCTs in patients with moderate to severe AD. The analysis was conducted in accordance with internationally recognised standards (e.g. EMA and Cochrane) and published in the peer-reviewed literature. Despite the meta-analysis being previously reviewed by EMA and also being available in the public domain, PenTAG excluded this from its review.
- Instead, only two studies in patients with moderately-severe to severe AD were reviewed by PenTAG, thereby excluding studies in patients with moderate AD and those on a stable dose of AChEIs. Justification for excluding the remaining four RCTs has no clinical basis and contradicts the approach taken by other independent review bodies.
- Given PenTAG's restricted review of the submitted evidence for memantine in patients with moderate to severe AD, conclusions drawn from the TAR regarding the clinical effectiveness of memantine are invalid.

5. PenTAG did not evaluate evidence submitted by Lundbeck in a sub-group of patients with behavioural disturbances, as defined by the final scope.

- The TAR acknowledges the impact that agitation/aggression and/or psychotic symptoms have on AD patients, their carers and healthcare providers.
- Lundbeck presented full and comprehensive analyses on the efficacy of memantine in this patient sub-group.
- A large body of evidence illustrating the significance these symptoms have on the clinical course of the disease and overall impact on AD burden was also presented.
- Despite evidence for high levels of unmet clinical need in these patients, the analyses submitted by Lundbeck for memantine in this behavioural sub-group were completely overlooked.

6. PenTAG did not review real-life observational data. Consequently, conclusions in the TAR for memantine are made without reference to its clinical effectiveness.

- In AD, as for many therapeutic areas, there is a need to consider outcomes that are pertinent to the real-life management of the disease and more removed from stringent clinical endpoints that characterise clinical trial design.

- Data on the real-life effectiveness of memantine was included in the Lundbeck submission. This was excluded by PenTAG, although in the TAR it is acknowledged that outcomes such as institutionalisation, carer time/burden and the use of antipsychotic medicine are not easily obtained from the RCT evidence alone.

PenTAG Approach to the MTA

7. PenTAG adopted an inconsistent and flawed methodology to their appraisal of the health technologies. Lundbeck have critical concerns about the approach to the evidence review and overall evaluation process taken in this MTA.

- The role of submitted unpublished evidence was unclear. The unpublished analyses that were performed to address the decision problem defined in the scope were omitted from the review. PenTAG acknowledges the importance of the behavioural domain in AD but completely overlooked the evidence submitted by Lundbeck in a sub-group with behavioural disturbances.
- The methods of reviewing and reporting consultee submissions were not transparent.
  - Each of the consultees' submissions was reviewed by different people within PenTAG and yet it is not evident whether a common check-list was used when appraising and reporting on the evidence. Discrepancies in the summaries across the appraised technologies may have arisen as a result of human error, rather than from major differences in the key quality control points (e.g. quality and relevance of the presented data or methodologies used).
  - The presentation of economic models developed by consultees, as reported in the TAR, was of a poor standard:
    - There was no acknowledgement of conservative assumptions and approaches used in consultees' models.
    - There was no acknowledgement of the strengths of consultees' models.
    - Disclosure of evidence by consultees to enable fair recommendations based on the best possible evidence was not recognised.
    - The TAR combined a description of the model with a critique of it, making it impossible for a lay reader to distinguish between the two, or to make a fair judgement on the strengths and limitations of the model.
    - The critique was unclear, inconclusive, poorly presented and lacked the necessary information to enable the consultee to initiate additional analyses in response. The critique gives the impression that the consultees' models are of poor quality, when it is PenTAG's description of the model that is questionable.
    - Most of the critique on Lundbeck's model was limited to a discussion of 'missing' details. However, this information was included in the Lundbeck submission, either in the main body of the document, the appendices or the references. Lundbeck provided clear citations in all cases.
    - Importantly, PenTAG did not acknowledge that many of the criticisms made on the consultee models, applied equally to the model PenTAG used.

- The methodology employed to appraise the clinical evidence generally lacked reliability:
  - The quality of memantine trials was misrepresented. The submitted trials are published in peer-reviewed journals and conformed to the CONSORT guidelines at the time of publication.
  - PenTAG provided an inappropriate description of excluded data and did not consider many of the data sources included in the Lundbeck submission.
  - The search strategy used by PenTAG did not identify at least one systematic review; a key systematic review for memantine undertaken by the Cochrane Collaboration and published in 2009 was not included.
  - PenTAG misinterpreted scales used to assess AD. In the MTC, the probability of memantine being most effective is lower than the AChEIs due to the incorrect interpretation of the global deterioration scale, in which a higher score indicates a worse health state.
  - PenTAG provided little justification for the methodology used:
    - While PenTAG recognised the SMD approach, it was not appropriately applied to the memantine data and thereby restricted a comprehensive review of all the available evidence.
    - Despite the known limitations of the LOCF in AD[3] PenTAG generally employed this method in their base-case analyses of RCT data.
    - The inclusion of memantine in the MTC analysis is debatable due to the differences in the baseline disease severity of patient populations in clinical trials for AChEIs and memantine. The results of such analyses have no clinical relevance. It is meaningless to compare the treatment effect of drugs in mild to moderate AD with those in moderate and severe AD, and inappropriate to use this method as a grounds for selecting the most 'effective' option. The assessment protocol specifies that treatment comparators should be in line with disease severity and licensed indications and therefore the MTC violates this protocol.

[3] Molnar FJ, Man-Son-Hing M, Hutton B, Fergusson DA.Have last-observation-carried-forward analyses caused us to favour more toxic dementia therapies over less toxic alternatives? A systematic review. Open Med. 2009 Mar 24;3(2):e31-50.

2 Cost-Effectiveness - Review of the PenTAG Model

The economic assessment of interventions is a key component in the overall review of the data, with the accepted incremental cost-effectiveness ratios (ICERs) a pivotal factor in the decision making process and final recommendations. Therefore, it is of vital importance that the assessment of cost-effectiveness is based on robust and well-validated models that:

- Reflect the decision problem
- Reproduce the relevant features of the disease and its management
- Take into account important patient characteristics and mimic the clinical course of the disease
- Make optimal use of available data with appropriate statistical methods
- Structure and implement the framework using suitable techniques and software.

These requirements are a minimum prerequisite for model credibility. In the case of Alzheimer's disease (AD), the use of an inappropriate economic model has serious consequences and can lead to many thousands of people being denied cost-effective interventions.

PenTAG developed a new economic model for the assessment of the cost-effectiveness of the acetylcholinesterase inhibitors (AChEIs) and memantine in the treatment of AD. A review of the PenTAG model and documentation reveals that the model does not meet the basic requirements for economic evaluation and includes several fundamental methodological flaws. The model fails both face and technical validity and thus does not provide reasonably accurate, meaningful or realistic estimates of the cost-effectiveness of AD treatments used in England and Wales.

Lundbeck believe the approach taken by PenTAG is distorted, particularly in relation to its applicability in assessing the economic case for memantine. The cost-effectiveness estimates of AD therapies in the UK clinical setting based on this model are therefore considered unreliable and should not be used to guide decisions about the appropriate use of NHS resources.

2.1 PenTAG Modelling Approach: Issues on Face Validity

2.1.1 Conceptual Framework of the Model

- **The important influence of cognition, functioning and behavioural symptom status on key outcomes in AD, such as disease progression and time to institutionalisation, is extensively documented. The inclusion of these three domains in any model of disease progression is well recognised, and was highlighted specifically in the TAR.**
- **Despite this, PenTAG excluded the behavioural domain from their model. The justification for this important omission is unclear but appears to be related to time constraints within which PenTAG were operating.**
- **The model therefore does not include an assessment of the benefit of treatment on the behavioural symptoms, and this has implications for the validity of the cost-effectiveness estimates generated.**
- **The model is not suitable for economic evaluations of therapies in moderate and severe AD, where behavioural symptoms, known drivers of institutionalisation, become prominent.**
- **The PenTAG model completely fails to reproduce key features of AD, its progression and its management.**
- **Peer-reviewed models that include all three domains are available in the public domain.**

In developing cost-effectiveness analyses for AD a key consideration is which clinical parameters most accurately predict the progression of disease. As PenTAG clearly acknowledges, while cognitive impairment is a major clinical feature of AD, on its own it is a poor predictor of disease severity, and particularly it is not highly associated with health-related quality of life or costs. AD is a complex disease with multiple factors influencing the patient's disease progression, including a patient's ability to perform day-to-day activities, behavioural disturbances and caregiver burden.

Although the true interaction between the different domains of AD and their impact on disease progression is unknown it is well-recognised that to undertake a comprehensive assessment of progression cognition, functional status and behavioural symptoms should all be considered.

PenTAG explicitly and unambiguously states the need to include the three domains of cognition, function and behaviour in order to undertake a robust analysis (section 7.2.3; page 260):

"*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"*

PenTAG indicate that cognition should be the first dimension included in the model and undertook a literature search to investigate whether either functional status or patient behaviour were valid and reliable independent predictors of quality of life or care costs. This led PenTAG to identify "*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*" (section 7.2.3; page 261).

PenTAG's recognition of the importance of behavioural symptoms on the progression of AD, and the need for inclusion of these symptoms to ensure a robust economic model are explicitly stated throughout the review.

However, despite this, the PenTAG model ignores behavioural symptoms and relies on the assumption that cognition, functional impairment and age are the only predictors of time to institutionalisation (see time to pre-institutionalisation equation on page 283 which incorporates age, MMSE [cognition] and Barthel ADL [function] only).

Failing to consider behavioural symptoms from the model is an important omission and the impact of this domain on relevant outcomes has been well documented in the literature. Indeed, in the background section of the TAR it states that behavioural symptoms "*have been shown to be better predictors of institutionalisation and carer distress than cognitive symptoms"* (section 2.2.5.2; page 46). Data from a longitudinal cohort study also presented in the TAR states that "*the severity of behavioural problems shown by the patient were predictors of institutionalisation odds ratio 1.08 (95% CI 1.01 to 1.15)*" (section 2.2.5.3; page 47).

The decision by PenTAG not to include behavioural symptoms within the model equation is not fully described and remains unclear. The TAR did state that "*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*" (section 7.1.1; page 254). Given the importance of behavioural symptoms in ensuring a robust model of disease progression, exclusion of these symptoms on the basis of time constraints is inappropriate and could lead to inaccurate cost-effectiveness estimates resulting in many thousands of AD patients in the UK being denied cost-effective treatments that have the potential to positively impact on their health.

Importantly, the UK-based study published by Wolstenholme and colleagues that was used to populate the model did collect data related to patients' behaviour ("*At four-monthly*

intervals the subjects were assessed in terms of their cognition and the carers were interviewed about the subjects' behaviour, ADL and all health, social and long-term care services used.", Wolstenholme et al., 2002, page 36). It should be noted that the investigators of the dataset used in the PenTAG model published an analysis of the same data and this included many additional predictors. The availability of this data is not discussed in the TAR.

The statistical model used by PenTAG was a multivariate regression model of time-to-institutionalisation. This is a standard exponential model that assumes a constant hazard of nursing home placement over time, representing the most simple parametric survival model. The use of lack of time as a reason to exclude the third dimension from the model seems particularly flawed as the behaviour symptoms could simply have been integrated as one additional covariate within the model. This would have addressed this limitation and would not have required extensive additional work. Lundbeck also have critical concerns on the adopted methodology and validity of the final model (this will be described in more detail in the next section).

An important implication of the approach to the predictive equation used in the PenTAG model is that it does not account for the treatment effect on one of the core AD domains, thereby producing unreliable estimates of cost-effectiveness.

It is also worth noting that the predictive equation utilised in the cost-effectiveness analysis of memantine submitted by Lundbeck was in fact based on a predictive equation that included all three dimensions (cognition, functioning and behavioural symptoms) defined as most important in the TAR. These predictors were selected using objective statistical considerations that revealed the three of them were significant independent predictors. The methodology for the development of this new predictive equation has been published in a peer-reviewed journal.[1]

2.1.2 Construction of Predictive Equation of Time to 'End of Pre-Institutionalisation'

- **The predictive equation used in the PenTAG model is methodologically unsound.**
- **The simplistic approach adopted by PenTAG does not allow for a consideration of the acceleration of hazard rates over time that has been observed in several evidence sources.**
- **In addition to age, the statistical approach relies solely on the uncertain and insignificant relationship between institutionalisation and cognition and functioning. This is in direct contradiction to other analyses on the same data set. Furthermore, the important and significant predictors identified in the original study analysis, such as gender, education, presence of behavioural symptoms and living arrangements, were ignored.**
- **The only significant predictor in the equation is age; a factor on which treatments can have no effect.**

Uncertain Relationship between Time to Institutionalisation and the Two Predictive Variables (Cognition and Functioning)

As previously described the PenTAG model is based on a predictive equation that models time to institutionalisation. Within the TAR the following descriptions of the equation are provided:

"*To calculate an equation representing time to end of pre-institutionalisation, an exponential survival regression model ("survreg" routine from the "survival" "R" package) was fitted, with time to end of pre-institutionalisation as the response variable, and MMSE, Barthel-ADL and age at the start of study as covariates.(…) For simplicity, the exponential distribution was chosen, rather than more complex two-parameter functions*" (section 7.3.8; page 282).

"*Although MMSE and Barthel-ADL were not identified as statistically significant variables in explaining the variance of time to end of pre-institutionalisation, both were retained in the model so that a treatment effect could be incorporated into the decision mode*" (section 7.3.8; page 283).

Lundbeck would like to raise several points in terms of the model used.

Firstly PenTAG utilised an exponential parametric survival model which relies on the assumption that the hazard rate is constant over time. The only justification provided to support the choice of exponential survival model is related to simplicity. There is clear evidence from previous work that the exponential approach does not properly fit. Time to full time care (FTC) and time to death were previously modelled by Caro and colleagues[2] who observed that both hazard rates were increasing with time. It is worth remembering that this study and the resulting predictive equations of time to FTC and death were the

basis of the model developed by the Southampton Technology Assessment Center (SHTAC) for the previous multiple technology assessment of treatment in AD in 2004. Also during the development of the Lundbeck model detailed investigations into the relationship between the hazard rate and time were conducted revealing that the hazard rates were increasing with time for both mortality and the transition to need for full time care.

Therefore multiple evidence sources tend to refute the PenTAG assumption that hazard rates of FTC and death are constant over time. Thus one can expect that hazard rates estimated by PenTAG will be over-estimated in the short-term period and then under-estimated in the long-term. The over-estimation of hazard rates in the short-term period is the most problematic as it implies under-estimated survival time and time to institutionalisation, which will then cascade into an under-estimation of the treatment benefits. The long-term period on the other hand is less problematic for several reasons (e.g. a smaller cohort due to attrition and reduction of outcomes dues to discounting on health benefits).

The survival function used in the PenTAG model fails to account for time-varying hazards and therefore, as all transitions between health states are determined by these inaccurate functions, the ability of the model to predict the disease course is questionable.

The second major limitation of the PenTAG predictive equation is the lack of statistical significance for the main predictors of the time to institutionalisation (MMSE and Barthel index).

The original investigators of the dataset used in the model found these factors to be significant predictors of time to institutionalisation and therefore it is very surprising that PenTAG elected to ignore the published data and undertake their own analysis, particularly given that they report time constraints as a limitation to their overall approach. The fact that the PenTAG analysis did not reproduce prior knowledge should have resulted in PenTAG reviewing their own analysis or contacting Wolstenholme and colleagues to understand the cause of the discrepancy.[3]

The table below compares the predictive coefficients from both the Wolstenholme and PenTAG regression models:

Predictors	Wolstenholme	PenTAG
Age	0.069025	-0.05735
MMSE	-0.100454	0.00409
Barthel index	-0.159344	0.02139
Other predictors	Gender, living situation, attitude of the carer, physical ability to cope with caring, wakefulness at night, aggressive behaviour	None

Comparison of the relative impact of the predictors on time to institutionalisation reveals critical discrepancies between these two models. While the impact of age on the risk of institutionalisation is comparable, 0.069025 and -0.05735 in Wolstenholme and PenTAG respectively, the weights that cognition and function have on the risk of institutionalisation are strikingly different in these two models. The Wolstenholme model showed that all three predictors are significant, with age bearing the least impact on risk of institutionalisation. The PenTAG model showed that age is the only significant predictors and its impact on risk of institutionalisation is the highest.

It is interesting to interpret the results of the models. In the Wolstenholme model, a difference of 0.7 (0.069/0.101) points on the MMSE scale would be needed to have the same influence on the risk of institution as a difference of one year of age if all other characteristics are the same. In the PenTAG model a difference on the MMSE scale as large as 14 (0.057/0.004) points would be needed to have the same influence on the risk of institution as a difference of one year of age. Notably, 14 points on the MMSE scale represents the difference between mild and severe AD. The same discrepancies between the models are observed on the Barthel index.

However, PenTAG did not address these contradictory results and instead incorporated cognition and functioning into the model by forcing them into the predictive equation. The fact that, in the PenTAG predictive equation, age is the only significant predictor of time to institutionalisation means that no treatments can have any significant effect as they clearly cannot impact on age. The non-significant nature of cognition and functioning in the model imply extremely small effects on these outcomes, making the high ICERs a foregone conclusion.

It should be noted that the inclusion of insignificant predictors is permitted when clinically justified, however significant predictors should remain in the model to allow for accurate predictions. This is not the case in the PenTAG model and therefore there is expected to be substantial variability in the estimation of the coefficients associated with these parameters. This uncertainty around the 'true' value of these coefficients may result in inaccurate predictions of time to institutionalisation.

Poor Reporting of the Predictive Equation

The statistical procedure undertaken to build the equation is poorly reported. Results of univariate analyses and the testing of the relationship between potential predictors and outcome (institutionalisation) were not presented. Furthermore, the selection process for the variables included in the model was not described. The true predictors of time to institutionalisation in the database and how they could potentially impact the final model when included in the equation along with insignificant predictors (MMSE, Barthel index) remains unclear.

It is also very important to note that the predictive equation utilised in the cost-effectiveness analysis of memantine submitted by Lundbeck was based on a predictive equation that included all three dimensions (cognition, functioning and behavioural symptoms) defined as most important by PenTAG. It is of note that only significant predictors were included in the predictive equation developed by Lundbeck. The model developed by Lundbeck would therefore appear to be more in line with the requirements for a robust model stated by PenTAG than the model they themselves developed. The methodology for the development of the new predictive equation including all three domains and used in the Lundbeck model has been published in a peer-reviewed journal.[1]

2.1.3 Modelled Population

- **The potential role of memantine in the treatment of AD is poorly considered as a result of the exclusion of key data reflecting sub sets of the target population.**

The PenTAG model based their statistical model of time to institutionalisation on an untreated patient cohort: "*Data regarding the characteristics of people with Alzheimer's disease were primarily based on IPD from the study by Wolstenholme and colleagues. 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*. "

The majority of AD patients have either been treated with AChEIs, are currently treated with AChEIs or are not treated with these therapies as a result of contra-indications. Within the UK the majority of patients with moderate and severe AD will receive AChEI at some point during their disease course. It is therefore deemed inappropriate to evaluate the cost-effectiveness of memantine in patients who are naïve to any AD medication as this is not a true representation of real-life practice or the target population.

The PenTAG model states that it models the impact of memantine in the treatment of moderate to severe AD patients. However, the data used in the model for memantine efficacy includes only patients with moderately severe to severe AD. Despite the availability of data for memantine in moderate to moderately severe AD this is not included in the model. Furthermore, the economic model does not consider patients currently on a stable dose of AChEIs.

The cost-effectiveness of memantine should be evaluated on the full body of evidence for the licensed indication and should include patients with moderate AD and patients on a stable dose with AChEIs.

2.1.4 Modelling of the Treatment Effect

- **The PenTAG model makes the assumption that no treatment effect is observed until six months. This is in contradiction to evidence from RCTs and has no clinical basis.**
- **The model relies on data from the study by Wolstenholme and colleagues, in which no standard functional disability scale was used. This necessitated a mapping procedure between the Barthel index score, the ADCS-ADL_{19}, and the ADCS-ADL_{23}.**
- **Although all scales measure functional domain, the Barthel index appraises only basic activities, while the ADCS-ADL scales also cover instrumental activities.**
- **The assumptions to compensate for items in the Barthel Index with no counterpart in ADCS-ADL led to an underestimation of these Barthel scale items, and thus of the total Barthel index score.**
- **This ultimately results in an underestimation of the true benefit of AD therapies and thereby an underestimation of the time to institutionalisation.**
- **It is of particular note that the memantine data was inappropriately handled. When assessing the effect of memantine function using the Barthel index, ADCS-ADL_{19} scores were directly used in the equation without conversion into the ADCS-ADL_{23} scores. Neglecting to convert ADCS-ADL_{19} to ADCS-ADL_{23} implicitly assumes that the score and benefits of treatments on the two scales are comparable, which is incorrect. This results in an additional underestimation of memantine benefit on functioning.**
- **Inaccuracies in the modelling of the treatment effect further invalidate the cost-effectiveness estimates from the PenTAG model.**

2.1.4.1 Onset of Treatment Effect

The PenTAG model assumes that there are no benefits of AD treatment from month 0 to month six as described here:

- *"For the initial treatment period (point A to B), mean time to institutionalisation 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 institutionalisation (point C) is predicted based on the mean baseline characteristics plus the mean treatment effect for the treated cohorts. (…)This leads to treated cohorts having a delay in institutionalisation 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"* (section 7.3.4; page 270).

- *"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."* (section 7.3.7.1; page 273).

The model therefore relies only on the length of clinical trials to determine the onset of the benefit of treatment, and thereby implicitly assumes no effect of the drug for the first six month period. A review of the clinical trials reveals this assumption to have no basis as data indicates that benefits are in fact seen much earlier than six months. This data is presented in the review of clinical effectiveness in the TAR, with data at 12 weeks presented for memantine. Although acknowledged as unrealistic by PenTAG ("*The incorporation of the full treatment effect at six months is artificial. It is more likely that improvements due to treatment are gradual.*", section 9.3, page 375), this assumption is still integrated into the PenTAG model.

Notably, the previous SHTAC model utilised a more realistic assumption: "Eligible patients start treatment immediately and benefits from treatment are assumed to have an immediate effect, modifying patients' time-related risk of progression from pre-full-time care to the full-time care health state" (Green et al., 2004).

The assumption underestimates the true benefit of the therapies within the model.

2.1.4.2 Mapping of Functional Scales

Underestimation of Treatment Effect

To account for the treatment effect on functioning for all drugs in the PenTAG model a mapping algorithm was use to convert Alzheimer disease cooperative study – activities of daily living (ADCS-ADL_{19}) scores into the Barthel ADL index. This is fully described on page 276 of the TAR but the approach can be broadly summarised as follows:

- Published observed data on the ADCS-ADL_{19} (ADCS-ADL severe) was collected
- A mapping approach was used to derive Barthel index scores from the ADCS-ADL_{19} data
- A mapping approach was used to derive ADCS-ADL_{23} scores from the ADCS-ADL_{19} data
- An equation was developed using regression techniques that related the ADCS-ADL_{23} to the Barthel index scores

One of the critical limitations of the approach lies in the discrepancies of activities of daily living (ADL) assessed by the three scales:

- The Barthel scale covers only basic ADLs (e.g. toileting, bathing, grooming) and none of the instrumental ADLs (e.g. making a meal, shopping, reading, writing)
- The ADCS-ADL_{19} scale covers basic ADLs and some of instrumental ADL
- The ADCS-ADL_{23} scale covers basic ADLs and instrumental ADL

This difference in terms of the content of the scales has implications on the mapping validity.

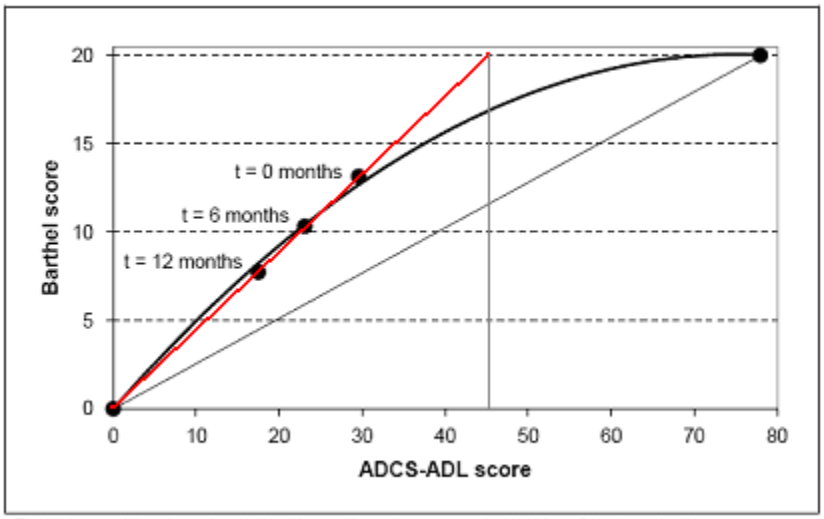
It is commonly known (see for instance Kurz et al., 2003)[4] that basic activities (e.g. feeding, mobility) are the most simple functions, and as such are kept relatively intact until the most advanced stages of the disease. On the other hand, instrumental activities (e.g. preparing a meal, shopping) start to deteriorate much earlier in the course of the disease. Therefore different scales are used to measure the functional deterioration at different stages of the disease (ADCS-ADL_{23} is used for mild and moderate AD, while Barthel and ADCS-ADL_{19} scales will be more suitable for advanced AD).

In the mapping approach undertaken by PenTAG, when no single question from the ADCS-ADL_{19} matched a given Barthel scale item, the rescaled total score on the ADCS-ADL_{19} (i.e. the sum of the basic and instrumental activities) was used as a proxy for the Barthel scale item. This implicitly assumes that the missing Barthel scale item ("missing" in the sense "no equivalent in ADCS-ADL_{19}") would have been an "average" item of the ADCS-ADL_{19}, if it had been measured in this scale. The term "average" refers to the ranking of ADLs, from the most simple (i.e. least impaired, with highest score) to the most difficult (i.e. most impaired, with lowest score). As the ADC-ADL_{19} is made of 6 basic activities (the first six items) and 13 instrumental activities, the above-mentioned missing Barthel scale item is then expected to be among the simplest ADLs and have a higher score compared to an "average" item of ADCS-ADL_{19}.

This obviously results in an underestimation of this missing Barthel scale item, which then cascades into an underestimation of the total Barthel index score. This ultimately leads to an underestimation of the treatment benefit estimated from this mapping approach and therefore an underestimation of the time to institutionalisation.

Also, as shown in figure 61, the relationship between the Barthel index and the ADCS-ADL score seems very close to linear if the constraint on maximum scores is omitted. Following a linear assumption, the maximum Barthel index score would then translate into a score below the maximum of the ADCS-ADL (between 45 and 50 on a scale of 78). This is consistent with the different contents of the scales considering that the Barthel index does not include instrumental activities of daily living. As instrumental ADLs are more complicated and deteriorate faster than the basic ADLs, a maximum score on the basic activities (Barthel index) does not guarantee a maximum score on all activities (ADCS-ADL).

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.

Building the predictive equations for time to institutionalisation and time to death using data from the Wolstenholme study, in which no standard functional disability assessment scale used in RCTs is used, makes the mapping between the Barthel index score and standard disability assessment scale unavoidable. This directly implies a longer chain of evidence to relate clinical efficacy to effectiveness and therefore a possible distortion of information between the two. This can unfortunately not be assessed as no observed data are available to validate or invalidate this approach.

Poor Reporting of the Mapping Algorithms

The methodology of the mapping procedures employed by PenTAG is poorly reported, and particularly insufficient information is provided to allow for replication of the exercise:

- The relationship between the Barthel scale items and the ADCS-ADL_{19} items (i.e. which item from the ADCS-ADL_{19} was used as a proxy for which item on the Barthel scale, and which items from Barthel scale had no single equivalent in ADCS-ADL_{19}) is not documented.
- The relationship between ADCS-ADL_{23} items and ADCS-ADL_{19} items is not documented.
- The "which most closely correlated" criterion used to match items from two different scales is not explicitly described. The "acknowledgement" section of the TAR tends to indicate this is more related to expert opinion than statistical correlation techniques, but no confirmation of this can be found in the report.

This lack of transparency regarding the mapping process thereby prevents the manufacturer from addressing properly the two above-mentioned points:

- Under-estimation of the effect of memantine resulting from assimilation of ADCS-ADL_{19} and ADCS-ADL_{23} scores (as detailed in the next section).
- Under-estimation of the effect of memantine resulting from use of instrumental ADLs to estimate scores on the Barthel basic ADLs items.

2.1.4.3 Inappropriate Handling of Memantine Data

As a result of the mapping exercise performed in the TAR, the following equation was derived relating the Barthel index and the ADCS-ADL_{23}:

Barthel score = 0.534 * (ADCS-ADL_{23} score) - 0.0036 * (ADCS-ADL_{23} score)^2

However, when assessing the effect of memantine function using the Barthel index, ADCS-ADL_{19} scores were directly used in this equation, without conversion into ADCS-ADL_{23} scores.. This approach is clearly inappropriate. Neglecting to convert ADCS-ADL_{19} in ADCS-ADL_{23} implicitly assumes scores and benefits of treatments on the two scales to be comparable, which is an incorrect assumption. The important differences between the ADCS-ADL_{19} and the ADCS-ADL_{23} are well acknowledged within the TAR as exemplified by the attempt to convert the ADCS-ADL_{19} data into ADCS-ADL_{23} data.

Despite the presence of a pool of common items (approximately 15) with identical scoring between the two scales, ADCS-ADL_{23} includes many additional activities of daily living, resulting in a larger range (0-78) compared to ADCS-ADL_{19} (0-54), implying larger scores and thereby the potential for a larger treatment difference when measured using the WMD. Assuming that the benefits of memantine treatment using the ADCS-ADL_{19} would be the same as that measured by ADCS-ADL_{23} is therefore a very unfavourable and non evidence-based assumption that results in an underestimation of memantine benefit on functioning.

It is worth noting that this incorrect handling of the data could only be identified from the electronic model and not from the TAR report, thereby highlighting the poor reporting by PenTAG.

2.1.5 Assumptions on Utility Value for Institutionalised Patients

- **The model makes two other unjustified assumptions that further negatively impact the validity of the cost-effectiveness assessment of memantine:**
  - **40% of moderate to severe patients (representing the licensed population for memantine) are in an institution at model start;**
  - **The utility for patients in an institution is equal to the utility in severe patients.**
- **These assumptions are based primarily on the basis of the licensed indication of AChEIs. The other clinical justification provided for this assumption is misleading.**
- **These inappropriate assumptions, while having minimal impact on the economic evaluation of the AChEIs, mean that the model is unable to show QALY gains due to prolonged time to institutionalisation in this population.**

In the PenTAG model the following assumptions are made about the proportion of patients who are in institutions at the start of the model "*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.*" (section 7.3.3; page 268).

The PenTAG model makes the assumption that institutionalisation is equivalent to the severe state of AD based on the following justification "*The PenTAG model allows for treatment discontinuations, and assumes that for the three cholinesterase inhibitors, treatment stops once they enter institutionalisation. Thus, the model implicitly assumes that institutionalisation 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)*" (section 7.3.4; page 272). The assumption that the institutionalisation is equivalent to a severe health state is not based on clinical data but rather solely on the indication of AChEIs and the fact that these therapies are stopped once patients are admitted to an institution. The use of this assumption makes the model invalid as it is not reflective of the true clinical situation in AD.

Furthermore, this assumption is in direct contradiction with the assumption that at model start 10% of patients with mild to moderate AD, the target population for AChEIs, are in an institution (i.e. considered as severe) and thereby being treated outside the indication for these therapies.

Another important point related to the assumption linking severe disease to institutionalisation is that 40% of moderate to severe patients (representing the licensed population for memantine) are in an institution at model start. The quality of life of all of these patients is assumed to be that of people with severe AD. This assumption means that there is no possibility for memantine to demonstrate any benefit on 40% of the target population within the model. Although explicitly acknowledged by PenTAG as major limitation ("*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", section 9.3, page 374), this assumption was still implemented in the PenTAG model.

PenTAG also makes the following statement in regard to the assumption on the utility of the institutionalised patients "*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 institutionalisation 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 institutionalisation*" (section 7.3.4; page 272).

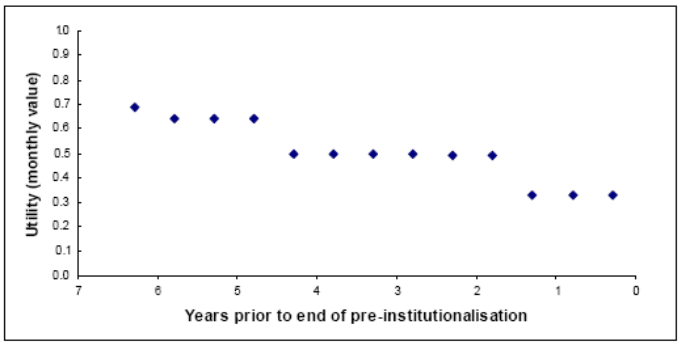
This interpretation based on mean time can be misleading. If it is assumed that the time to MMSE score of 9 at the time of institutionalisation follows an approximately symmetric distribution this means that at the time of patients are institutionalised approximately half have an MMSE below 9 (representing the severe patients) but that importantly the other half have an MMSE above 9 at this time point. This therefore does not establish any concordance between severity and institutionalisation.

The inappropriate assumption that the utility for patients in an institution is equal to the utility in severe patients has only a minimal impact on the economic evaluation of the AChEIs as treatment with these agents is stopped once patients reach the severe AD state or enter an institution. However, the assumption has important and negative consequences for the evaluation of memantine as it implicitly assumes that any treatment effect that memantine has in a severe patient will never translate in QALY gains by prolonging to time to institution.

It is noteworthy that PenTAG wanted to explore "*gradually decreasing health-related quality of life in the time before patients become institutionalised*" (page 263), yet the assumption they used makes this unfeasible.

It is evident from figure 68 in the TAR (reproduced below) that the utility of patients is estimated to remain constant for one and a half years prior to institutionalisation. As a consequence, the ability of the model to measure the benefit of a drug in severe AD is questionable, especially in a population with the more advanced stages of disease who are rapidly moving towards needing institutionalisation. This population are likely to be treated with memantine.

FIGURE 68 *Plot of utility from Jonsson and colleagues[201] by time to end of pre-institutionalization used in the base-case analysis*



2.1.6 Estimating Cost of Treatment and Patient Management

- **There are a number of assumptions within the model that could result in the costs for memantine being over estimated.**

In the model, as described above it is assumed that AChEIs are discontinued at the point of institutionalisation. However, PenTAG state the following "*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 memantine is assumed to be taken by individuals until they die*" (section 7.3.4; page 272). This assumption is inappropriate and may over estimate the cost of memantine as it is unrealistic to assume that patients would continue on memantine treatment for this length of time. As described in the Lundbeck submission the typical duration of memantine treatment in the UK is 16 months.

Another inappropriate assumption related to the cost estimates is as follows "*They are the drug costs, monthly costs of care (pre-institutionalised and institutionalised) and the costs of a 6-monthly monitoring outpatient visit for those treated with donepezil, galantamine, rivastigmine or memantine.*" (section 7.3.10; page 299). In the PenTAG model, although not explicitly stated here the cost of the monitoring visit is assumed to be the cost of a specialist visit. This is a valid assumption for treatment initiation, however for prescription renewals only standard GP visits are necessary. As the cost of GP visit is much lower than the cost of a specialist visit (£36 versus £124.28), this represents an overestimation of the monitoring cost.

A final comment on the estimation of drug costs relates to the following "*Monthly inflated cost (£) = 3363 - 1117t + 191t^2 – 10.7t^3.*" (section 7.3.10; page 302). "*In the cost–utility model, at each cycle, the proportion of the cohort within 6-monthly time-periods of leaving the pre-institutionalised 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 institutionalisation." (section 7.3.10; page 304). It is not clear why, after having computed an equation for the cost at any time-point according to time remaining in pre-institutionalisation a time-interval approach as opposed to using the exact time is employed. No justification for this is provided.

2.2 Implementation of the PenTAG Model: Issues on Technical Validity

- **In addition to the issues of face validity described in section 2.1 above, there are several issues in terms of the technical validity of the model.**
- **Given the complete lack of face validity, it is somewhat moot to consider the technical validity (i.e. does the model compute correctly). Nevertheless, the PenTAG model also fails in this regard.**
- **The implementation of the PenTAG model does not reflect the model concept. For example, the model does not represent a Markov structure as there were no transitions between the four states ("pre-FTC/inst", "FTC/inst", "alive", and "dead").**
- **PenTAG implemented a modelling structure that purports to allow for diminishing utility and rising costs over time thereby addressing limitations of the previous model. However, an inspection of the model reveals that the calculations do not implement these changes and that the calculations used in fact generate values that vary considerably from the source data. It should also be noted that in the model utilities vary with age although no such relationship was postulated.**
- **The model does not behave correctly as exemplified by the fact that small changes in mortality lead to negative numbers in the 'inst' case leading to the total costs being lower than the pre-institutionalisation costs.**

2.2.1 Does Implementation Reflect the Model Concept?

The concept of the PenTAG model was very straightforward: three states with transitions from the pre-institutionalisation to the institutionalised state, and from both to the dead state. Despite this extremely simple concept of a complex disease, PenTAG did not implement it as specified. In the actual model there are four columns representing the structure labelled: "pre-FTC/inst", "FTC/inst", "alive", and "dead". Inspection of the formulae reveals that there are no transitions among the states and that the model is not a Markov structure. Instead, the model is of the survival type. A failure-time equation is applied in the column "pre-FTC/inst" which decreases the proportion of the cohort in this state. A separate survival equation is applied in the column "alive" to determine the proportion of the cohort still alive. The proportion in the "FTC/inst" column is determined by subtracting the proportion in "pre-FTC/inst" from "alive". The proportion "dead' is simply the complement of "alive". This structure not only does not match the model concept but leads to errors as the

application of the two failure-time equations is not linked in any manner, as documented below.

One of the shortcomings of the previous SHTAC model was the lack of any deterioration in quality of life before reaching FTC (or of any increase in costs), despite the disease progression over time. To address this shortcoming, PenTAG implemented a separate modelling structure that purports to allow for diminishing utility and rising costs. This structure attempts to compute the monthly distribution of time to institutionalisation in the pre-institution state and use this to derive a weighted average utility and cost to apply each month. However, a simple examination of this structure reveals that it is exactly the same every month, contradicting the idea that gradations in utility and costs were being implemented. The distribution only begins to change when the secondary model runs out patients in the longest time-to-institutionalisation strata, forcing a change in the distribution (although by this time the entire cohort is dead). Thus, notwithstanding what the TAR says there is no implementation of a decreasing utility or of increasing costs. Moreover, the calculations lead to the use of values that do not correspond to the source data, as noted below.

2.2.2 Computation of Formulae and Use of Software

The secondary structure used to compute the distribution of time-to-institutionalisation for weighting of utilities and costs presents two problems. First, it references rows that are well past the end of the model time horizon. Secondly, and more importantly, it implies utilities that depart considerably from the source data, do not change over time as noted above, but change with age (although no such relationship was postulated). For patients with mild-to-moderate disease, the secondary structure is computing a constant utility of 0.5 in the oldest age group while the input data would imply 0.6. For patients with moderate-to-severe disease, the model uses 0.4 compared to an input value of 0.5. Utilities used in the model drop by 0.04 units with increasing age strata in mild-to-moderate disease (0.58 to 0.54 to 0.50) and by somewhat more than that in moderate-to-severe (0.51 to 0.46 to 0.42). Similar problems occur with the costs.

The secondary structure also introduces an inconsistency in the key input – the starting MMSE. In moderate-to-severe disease, for example, the weighting implies a starting MMSE score of 16.3, 13.1, and 10.3 (for the age strata), instead of the age-invariant 11.73 which should have been used. In mild-to-moderate, the discrepancy is even worse: 16.6, 13.7 and 10.8, instead of the supposed value of 17.1.

Apart from these major errors, the model is built inefficiently, with numerous repetitions of the same calculation when once would have sufficed; and with multiple echoing of the same value, rather than referencing the original one. This not only slows execution of the model, it makes the model less transparent and more prone to error.

2.2.3 Model Testing

As can be deduced from the above, the model does not behave correctly when values are changed. For example, the proportion of the cohort in the "pre-institution" state is completely immune to any changes in the probability of death — no matter how mortality changes, the proportion of patients over time who are pre-institutionalisation remains the same. This implies that all deaths come from the "institution" state, and is completely contrary to the model concept. More problematically, small changes in mortality well within the bounds of the inputs, lead to **negative** numbers in the "institution" state and these negative proportions of the cohort are multiplied by the utilities and costs and accumulated into the results. This leads to the bizarre results that total costs become lower than the component pre-institutionalisation costs and also to negative QALYs accumulating in the "institution" state.

Another clear indication that the model is not calculating correctly can be seen when the proportion of patients discontinuing treatment is changed: the higher the proportion, the better the cost-effectiveness. This illogical result occurs because discontinuation only affects the costs but not the effectiveness. Thus, the benefit is retained even if everyone stops before the end of the treatment period. The fact that memantine is dominant if patients are treated for less than the six months and then discontinued highlights the inconsistence of the model.

3 Cost-Effectiveness - PenTAG Review of the Lundbeck Model

The economic evaluation conducted by Lundbeck adopted a robust and well calibrated model to evaluating the cost-effectiveness of memantine in its target population:

- The Lundbeck model adopted the framework developed for the previous technology appraisal in 2004, however with improvements to address the key limitations of the SHTAC-AHEAD model. A detailed comparative analysis of the Lundbeck model and the SHTAC-AHEAD model is provided in Appendix A.
- In contrast to the PenTAG model, the Lundbeck model uses a predictive equation that includes all three domains (cognition, functioning and behaviour) identified by the TAR as being necessary for any comprehensive economic evaluation in AD.
- The Lundbeck model was built using data from the LASER-AD database, which represents a cohort of AD patients from the UK that best reflects routine clinical practice. These patients are also exactly aligned with the licensed population for memantine.
- Both the predictive equation used in the Lundbeck model and the use of the model in assessing the cost-effectiveness of memantine in the UK setting are publically available in the peer-reviewed literature[4,5].
- There were no major critiques on the model submitted by Lundbeck that would result in a change to the model conclusions or necessitate extensive additional analyses or structural changes to the model.
- The vast majority of the comments from PenTAG were related to 'missing' information, which could in fact be found in the submission, accompanying appendices or the references provided.

PenTAG made a number of criticisms of the Lundbeck model. However, the PenTAG model includes many of the same limitations that were raised as criticisms of the Lundbeck economic model. These are detailed in Appendix B. The vast majority of the comments from PenTAG were related to so-called 'missing' information, which could in fact be found in the Lundbeck submission, accompanying appendices or in the references provided. In addition to the major themes above the full responses to this individual critiques of the Lundbeck model are provided in Appendix C.

We believe that the Lundbeck model is a robust, validated and well-calibrated model for the assessment of the cost-effectiveness of memantine. The fundamental concerns with the PenTAG model and specifically the concerns about its applicability to the memantine case have already been described fully in section 2.

[4] Rive B, Le Reun C, Grishchenko M, et al., 2010a. Predicting time to full-time care in AD: a new model. *JME*, 13(2): 362-70.
[5] Rive B, Grishchenko M, Guilhaume-Goulant C, et al., 2010b. Cost-effectiveness of Memantine in Alzheimer's Disease in the UK. JME, 13(2): 371-80.

Lundbeck believe that substantial work must be undertaken before the new PenTAG model is fit-for-purpose and provides estimates of the economic efficiency of the AD treatments that are reasonably accurate, meaningful and realistic enough to permit guidance on decisions for access to treatments to be made.

It is therefore recommended that the Lundbeck model represents a much more appropriate and methodologically sound approach to the fair and clinically relevant economic evaluation of memantine. We propose that on this occasion the two different classes of drug in AD are assessed using different models.

4 Clinical Effectiveness

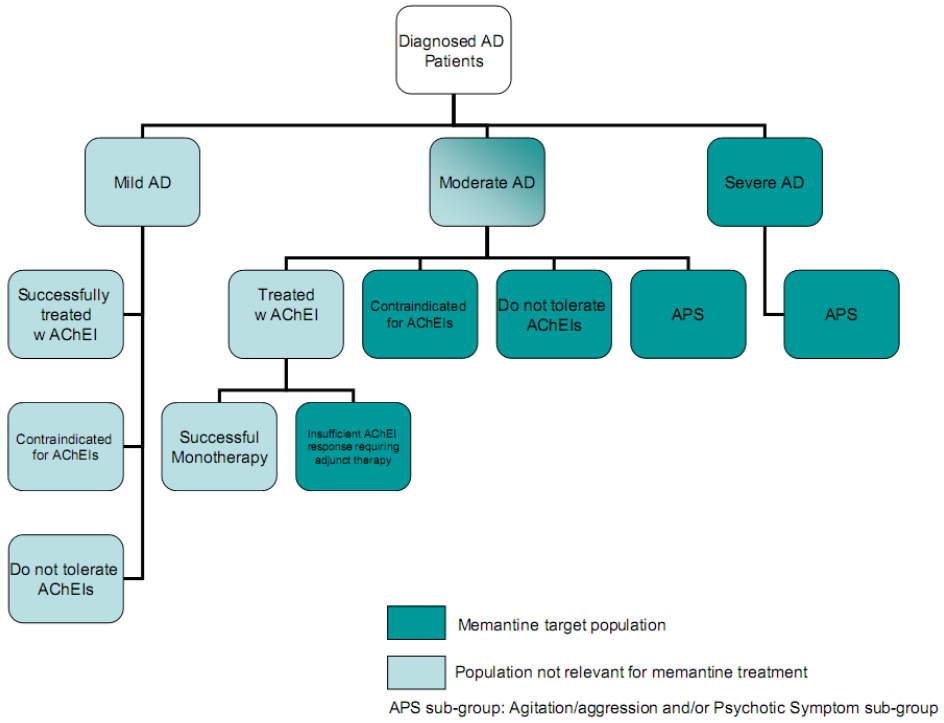
Memantine is approved for the treatment of moderate to severe AD (MMSE ≤19). Memantine can be used in:

- Patients naive to AD medication
- Patients who have a history of past use of AChEIs
- Patients currently on a stable dose of AChEIs

The efficacy of memantine in this population has been comprehensively demonstrated through evidence from randomised controlled trials (RCTs). The benefit of memantine in clinical trials has been substantiated through effectiveness evidence available from studies that represent routine clinical practice. In addition the efficacy of memantine in patients with agitation/aggression and/or psychotic symptoms (APS) has been comprehensively demonstrated in sub-group analyses.

Given the established treatment pathways for AD in the UK, memantine is best positioned to be used in the following patient populations (Figure 1). This target population represents a vulnerable patient group that currently have very limited treatment options and a high unmet need.

Figure 1: Memantine target population



The PenTAG review of the clinical effectiveness data for memantine is associated with several limitations. Several of these limitations relate specifically to the inputs used in the PenTAG economic model. Furthermore, the omission of key data means that the efficacy of memantine in the target population cannot be appropriately assessed.

4.1 Review of Memantine in the Licensed Indication

- **The remit of PenTAG was to review the clinical effectiveness of the AD therapies within their licensed indications. For memantine this is moderate to severe AD. This includes patients who are naive to AD medication, or have a history of past use of AChEIs or who are currently on a stable dose of AChEIs.**
- **However, PenTAG limited their assessment of memantine to moderately-severe to severe AD patients only. Furthermore, their review excluded patients who were on a stable dose of AChEIs. These exclusions were made despite the availability of robust data that has been published and used by the EMA to appraise memantine in its licensed indications.**
- **The fact that PenTAG undertook a restricted review of the evidence was not clearly stated.**
- **The exclusion of patients with moderate to moderately severe AD and patients on a stable dose of AChEIs, without a valid clinical justification, represents a major limitation of this review and renders the TAR conclusions on the efficacy of memantine invalid.**

In 2002 memantine received European approval for the treatment of moderately-severe to severe AD (MMSE $\leq$ 14). This indication was extended in 2005 to include moderate to moderately severe patients. The basis for this license extension was a meta-analysis of six large, phase III, placebo-controlled RCTs.

The meta-analysis was conducted in accordance with international standards for undertaking such analyses (e.g. EMA and Cochrane collaboration). Additionally, as described below the methods for selecting the specific AD patient populations for inclusion in the meta-analysis were in line with a published Cochrane review for memantine in AD and the previous TAR conducted in 2004.

The six trials included in the meta-analysis were considered pivotal by the EMA and were the basis for their conclusion that memantine has a statistically significant effect in moderate to severe patients. This meta-analysis is available in the public domain with data across key outcomes published in several peer-reviewed journals:

- Winblad 2007 – All domains[5]
- Mecocci 2009 – Cognition domain[6]
- Gauthier 2007 – Behavioural domain[7]
- Winblad 2010 – Functional domain[8]

In the meta-analysis three of the six RCTs were conducted in patients with moderately-severe to severe disease (Mini-Mental State Examination [MMSE scores ≤14]). Three others were performed in patients with mild to moderate disease (MMSE scores ≥10). All moderate to severe patients were included in the meta-analysis (N=1826). The trials include assessments of memantine in patients who were naive to AD medication, in patients who have a history of past use of AChEIs and in patients on a stable dose of AChEI. The

population included in the meta-analysis therefore reflects the licensed indication for memantine and also the memantine target population in UK clinical practice as described in Figure 1.

The Lundbeck submission dossier included a full description of each of the six trials included in the meta-analysis and the results of the meta-analysis across the key domains.

The aim of the TAR was to review the effectiveness and cost-effectiveness of memantine within the licensed indication of moderate to severe AD (MMSE ≤19). However, despite the availability of the comprehensive evidence base in this population PenTAG did not use the meta-analysis and included only two of the six RCTs within their appraisal of the efficacy of memantine. PenTAG inappropriately excluded key data for memantine in:

- Patients with moderate AD
- Patients on a stable dose of AChEI

Exclusion of Patients with Moderate AD

The justification of exclusion of data for memantine in patients with moderate AD was that "*individual patient data (IPD) was required to enable this sub-group analysis*" (section 4.4.3; page 76). However, data from individual trials for the moderate to severe population is available in the public domain as forest plots included in the meta-analysis publication. Furthermore, full details of this meta-analysis were provided within the Lundbeck submission. PenTAG did not identify the meta-analysis in their review of the data and despite this full data set being publically available and provided to them they did not include this data thereby excluding the moderate to moderately severe patient population. PenTAG was not explicit that the data related to the restricted patient population and excluded a key patient group. Exclusion of these patients from the analysis overestimates the uncertainty around the clinical efficacy of memantine in the licensed indication of moderate to severe AD.

Exclusion of Patients on Stable AChEIs

The exclusion of evidence for memantine in patients on a stable dose of AChEIs was justified on the following basis "*In the 2004 review, this* (the efficacy of memantine vs. placebo in patients on a stable dose of donepezil) *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*" (section 4.8; page 173).

As explored further below there is no clinical basis for the separation of these regimens and the decision of PenTAG to categorise the data in this way is inappropriate for several reasons.

Firstly, the RCT excluded by PenTAG for the main review of memantine effectiveness (but partially considered in their separate review of combination regimens) was designed to assess the efficacy of memantine compared to placebo in patients on a stable dose of AChEI. This is a highly relevant clinical question within the context of current clinical practice in the UK where the majority of patients who are eligible for treatment with memantine

have previously been treated with AChEIs or most importantly in this case are currently treated with a stable dose of AChEIs. If the aim of the trial had been to assess the efficacy of memantine as a combination treatment four arms (placebo only, memantine only, AChEI only, memantine + AChEI) would have been needed. The classification of this trial as assessing a combination regimen is therefore incorrect.

Exclusion of Patient Groups Contradicts Ongoing Clinical Trials

The inappropriateness of splitting these populations is further highlighted by considering the clinical trial designs for emerging therapies in AD. There are generally no strict inclusion/exclusion criteria in regard to prior or concurrent pharmacological treatment with included patients representing a mixed population of treatment naive patients, patients with prior AD therapy use and patients on stable AChEI or memantine use (Appendix D). This is clearly exemplified through a consideration of the inclusion and exclusion criteria for ongoing clinical trials of new AD medications. These include bapineuzumab (expected market approval for AD between 2012-2014), gammagard (expected market approval for AD in 2012), dimebon (expected market approval for AD in 2014+). The protocols and clinical trial designs for these new agents have been reviewed in collaboration with regulatory authorities; "The protocols for the four Phase 3 trials for bapineuzumab, which are approved by regulatory authorities […]" [9] Medivation received regulatory guidance from the FDA on pivotal trials with dimebon in AD.[10]

In fact, the EMA highlight that "*In many countries symptomatic treatment of dementia with cholinesterase-inhibitors is considered as standard of care, particularly in mild to moderate Alzheimer's disease. Therefore in the future new treatments for dementia may be evaluated more and more by using add-on-designs, particularly in long term studies the "pure" use of placebo control for demonstration of efficacy may be difficult to justify.*"[11]

Interaction Analyses

Thirdly, given the diversity of the patient populations included in the meta-analysis for memantine a number of interaction analyses were undertaken to examine the impact of baseline characteristics (including past or concurrent stable use of AChEI) on the efficacy of memantine. There was no significant interaction between treatment effect and these baseline factors. This validates the approach of considering these studies together. This also supports the assumption of an additive effect of memantine when used in patients receiving stable doses of AChEIs, thereby addressing this critique raised by PenTAG.

Therefore the combination of trials for memantine across these populations, as conducted in the published meta-analysis, is clearly an appropriate approach which provides an assessment of memantine in the licensed population that represents clinical practice. This is most clearly highlighted by the fact that this approach follows international standards, being used both by the EMA in their assessment of memantine and by the Cochrane collaboration in their 2009 systematic review.[12] In fact, this approach was also used in the previous Technology Appraisal conducted in 2004.

Data for Use in Meta-Analysis

A further point for consideration is how the data is reported for use in meta-analysis. Within the TAR data was primarily reported as weighted mean differences (WMD). This is an

absolute measure of effect size and is an appropriate measure in cases where the data being pooled across trials is all collected using the same scales.

Within the assessment of AD, given the large variation in patient characteristics depending on the severity of disease, it is not possible to use a single instrument to assess each domain across all patient groups. For example, when assessing cognition the Alzheimer's disease assessment scale – cognitive subscale (ADAS-cog) is the standard instrument for mild to moderate patients, whereas the severe impairment battery (SIB) is considered most appropriate in moderately severe to severe patients.[13] In the meta-analysis for memantine the severity of the included patients varied across trials and as would be expected a range of instruments were therefore used. In such cases, the standardised mean difference (SMD) should be used in the meta-analysis as this allows for assessment of the same outcome even when this has been measured in a variety of ways. This approach is well recognised and was employed where appropriate in the Cochrane systematic review of memantine.[12] It is of note that PenTAG raise the issue of the need for standard AD scales throughout the submission, for example "*Such good quality trials should aim to use the same standardised measures of cognitive status, functional status/ADL, and behavioural/psychiatric symptoms"* (section 1.6.2, page 39). This represents a lack of knowledge about the disease, as the use of the same scales in all severities cannot be justified as some scales become insensitive to detecting changes with increases in disease severity.

The TAR does recognise the SMD approach but states "*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*" (section 4.1.5.1, page 67). This approach is contradictory to well recognised approaches to meta-analysis and restricted some of the analyses that PenTAG could conduct on the data for memantine.

Finally, it is important to note that the TAR states that despite the differences in approach to the included data "*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*" (section 4.4, page 76). This statement is misleading as, due to the different data selected, the conclusions of the two evaluations are clearly different.

4.2 Interpretation of the Evidence for Patients on a Stable Dose of AChEIs

- **Study MEM-MD-12: Porsteinsson et al, 2008 includes mild AD patients who fall outside the memantine indication. This is not clearly stated in the TAR. The analyses should be based on the licensed indication for memantine, as stipulated in the methods section of TAR.**
- **The significant efficacy of memantine on all domains in MEM-MD-02 should be highlighted.**

- **Differences in the baseline severity of the patients from MEM-MD-02 and MEM-MD-12 should be explained in more detail and provided as a possible reason for the differences in clinical outcomes.**

The TAR considered the evidence for memantine in patients treated with a stable dose of AChEIs as a distinct group. As described above this is an inappropriate distinction with no clinical basis and is not in line with the approach taken by the EMA and Cochrane collaboration.

Even within the review of these trials there are several points of note.

Firstly, it should be noted that in the TAR the included trial for memantine in patients on stable AChEIs was MEM-MD-12: Porsteinsson et al, 2008. This trial is in a mild to moderate population. No justification is provided to explain why this study is included and it is not clearly stated that mild patients are considered. These patients fall outside the licensed indication for memantine. The inclusion of this trial for the PenTAG 'combination analysis' is therefore inconsistent with the exclusion of the trials of memantine as a monotherapy using the excuse that they are in mild-moderate or the other study with prior AChEI use conducted in the moderately severe to severe population (MEM-MD-02).

The lack of significant benefit in study MEM-MD-12 is highlighted although no reference is made to the significant efficacy that was reported in MEM-MD-02 across all the domains. This omission is particularly important as MEM-MD-02 is the only trial for adjunctive use of memantine that includes exclusively patients within the licensed indication for memantine (i.e. no mild patients). Of note, PenTAG states that for the cognition outcomes "*nor would it have been informative to combine two RCTs on a standardised scale*." although combination using the SMD approach is recommended by Cochrane in such instances.

Within the TAR, for both cognition (section 4.8.1.2.1; page 176) and functional outcomes (section 4.8.1.2.2; page 176) it is stated that the new study MEM-MD-12 cannot be pooled with the 2004 study MEM-MD-02 due to the use of different scales. However, as previously described the use of different scales in trials of AD patients with varying severities is well recognised and the SMD is a valid method for combining data in these cases.[12] Results using this method demonstrate significant benefit on the two outcomes.

The data from the two studies (MEM-MD-12 and MEM-MD-02) are combined only for the behavioural and global outcomes, with no such significant benefit observed despite significant results in MEM-MD-02 and PenTAG conclude "*no overall benefit from combination therapy"* (section 4.8.1.2.4; page 179).

The TAR highlights the difference between the results stating "*However, it is unclear why the behavioural and global outcome results are different*" (section 4.8.1.4; page 180). The TAR proposes two possibilities. The first is based on the concomitant medication used; "*The designs of these studies differed in that Porsteinsson and colleagues* [MEM-MD-12] *combined memantine with any of the three included AChEIs while Tariot and colleagues only combined memantine with donepezil* (MEM MD-02}" (section 4.8.1.4; page 180). This is also presented in the summary "*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" (section 1.3.2; page 29). However, these statements are misleading as patients in the MEM MD-12 study were predominantly treated with donepezil (over 65% of included patients) with only a minority of patients treated with galantamine (~15%).

Secondly PenTAG propose that the definition of ITT impacts the results;"*The other notable difference is that the 2004 authors analysed a modified ITT population whilst the 2008 study authors analysed a full ITT population*." (section 4.8.1.4; page 180). This is incorrect as both studies use the same criteria to define the population analysed (All randomised patients who took at least one dose of investigational medicinal product (memantine or placebo) and had at least one valid post-baseline assessment on primary endpoint).

The difference in the baseline severity of the included patient populations between the two studies is not included as a possible explanation for the variation in the results. This is an important consideration and should be included as a possible explanation for the observed differences.

4.3 Exclusion of Evidence in the APS Sub-Group

- **The efficacy of memantine has been demonstrated in a patient population with APS.**
- **While PenTAG recognises the burden of these symptoms on patients and carers, and their impact in predicting institutionalisation and decline, the data for memantine in this patient population was not included in the review as a result of incorrect exclusion criteria being applied.**

The background section of the TAR provides a detailed description on AD, including the symptom burden for the patient. This includes recognition that "*Commonly there are also neuropsychiatric symptoms such as anxiety, wandering, irritability, disinhibition and apathy*" (section 2.2.5.1; page 45). The impact of these symptoms on the carer and on the need for institutionalisation is recognised "*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 and carer distress*" (section 2.2.5.1; page 46).

Within the Lundbeck submission it was highlighted that the sub-group with agitation/aggression and/or psychotic symptoms (APS) has been consistently shown to differ from the remaining population in terms of patient characteristics (greater clinical decline), medical need (frequent misuse of antipsychotic drugs in AD) and higher economic burden.[14] [15] [16] [28] An analysis of memantine specifically in this patient population who have a greater unmet medical needs was therefore conducted and demonstrated a significant effect of the treatment in this population.

However PenTAG did not specifically consider this subgroup. They stated "*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" (section 4.4.3; page 77)

Individual patient data was required to perform analyses in the APS group. The analyses in the APS group were fully documented in the Lundbeck submission and all the information needed by the reviewer to make a well informed and fair judgement on the quality and relevance of this data was included.

4.4 Exclusion of Evidence on Clinical Effectiveness from Observational Studies

- **Despite the availability of a wealth of clinically relevant data from observational studies highlighting the effectiveness of memantine, PenTAG focused on systematic reviews and RCTs only.**
- **A discussion on the real-life effectiveness data for memantine within the context of the overall burden of AD should be included.**

In addition to the controlled clinical trials for memantine the submission from Lundbeck included a number of studies that present data on the real-life effectiveness of memantine. This includes the impact of memantine on lowering antipsychotics consumption, delaying time to institutionalisation, and on long-term cognition and functioning.

In the background section of the TAR the impact of AD on patients and carers is described, as well as the economic burden of the disease. Given the broad impact of AD across a multiple range of stakeholders there is a need to consider outcomes that fall outside stringent clinical outcomes captured in clinical trials and that are pertinent to the real-life management of AD (such as time to needing full time care or institutionalisation). However, PenTAG did not include any of this data in their review, despite their remit being to assess the effectiveness of AD treatments. It is of note that the PenTAG themselves recognise the limitations of the RCT evidence "*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 anti-psychotics*" (section 1.5; page 34) and the need for non-RCT evidence ("*Systematic reviews of non-RCT evidence on the impact of anti-AD treatments on resource use, institutionalisation and mortality*", section 10.2 page 379) that is identified as a research priority.

Reference is made to the discrepancies in the data considered between their review and that conducted by Lundbeck *"The reasons why some studies were included in the Lundbeck analysis but not included in the PenTAG meta-analysis are documented in Appendix 10"* (section 4.4.3; page 76).

In the protocol PenTAG states that systematic reviews and RCTs are the preferred data sources. However*, "the review protocol made a provision for broadening search criteria to include some observational evidence if insufficient systematic reviews or RCTs were identified"* (section 4.1.2.1, page 62). In the analysis of memantine as a monotherapy treatment PenTAG state that the conclusions are based on "*two moderate to poor quality*

trials and may be untrustworthy" (section 1.3.2, page 30). Given the recognition of the low number of RCTs available for memantine (resulting from inappropriate selection of studies by PenTAG), and that there is extensive additional and highly clinically relevant observational data available, the exclusion of this data is inappropriate

4.5 Exclusion of Key Safety Data

- **The development of conclusions on safety based on a single trial is inappropriate when there is data publically available that considers a pooled safety analysis.**

In the TAR data on the safety for memantine as a monotherapy is reported based on one study only MEM-MD-12 (in which patients are treated with stable dose of AChEIs) and reports that "*the main AEs in the memantine group were agitation and hypertension*" (section 4.6.4.3.6; page 153). It should be noted that the incidence of agitation was lower in memantine treated patients than placebo treated patients.

In the case of safety it is more appropriate to synthesise data across multiple trials. A meta-analysis on the tolerability and safety data from clinical trials published in 2008 would be a more appropriate source of safety data.[17] Other reviews of safety data from all memantine clinical trials are also available. This analysis reports that the most common adverse events with memantine are agitation and falls and both have numerically lower incidence than placebo.

4.6 Methodological Considerations

- **The quality of the memantine trials was misrepresented by PenTAG.**
- **The meta-analysis submitted by Lundbeck was conducted in line with internationally recognised standards. The inclusion of trials was clearly described.**
- **The LOCF method of analysis, preferred by PenTAG group, may be inappropriate in the assessment of AD due to the progressive nature of the disease. Analyses based on observed cases (OC) should also be considered**
- **The SMD represents a valid approach for pooling data for memantine when multiple scales have been used for a single outcome and these analyses should be included in the TAR.**
- **The inclusion of memantine in the MTC analysis is debatable due to the differences in the baseline disease severity of patient populations in clinical trials for AChEIs and memantine. The results of such analyses have no clinical relevance. It is meaningless to compare the treatment effect of drugs in mild to moderate AD with those in moderate and severe AD, and inappropriate to use this method as a grounds for selecting the most 'effective' option. The assessment protocol specifies that treatment comparators should be in line**

with disease severity and licensed indications and therefore the MTC violates this protocol.

4.6.1 Quality of the Memantine Trials

In relation to the trials found by PenTAG's search and included in their meta-analyses (MRZ-9605, , MEM-MD-02, MEM-MD-12 and MEM-MD-01), the TAR states that '*taken as a whole the quality of the trials was disappointing….'.* In relation to the memantine trials the TAR also states that the '*quality of the trials was moderate to poor…. "*(section 1.3.2, page 30).

The TAR appears to confuse the issue of quality of the trials with the quality of the reporting of the trials. The TAR fails to distinguish between the two and transparency is required to address these aspects separately.

Quality of the memantine trials

For trials MRZ-9605, MEM-MD-02, MEM-MD-12 and MEM-MD-01 (included in the TAR) plus MEM-MD-10 and Lu-99679 (the two additional trials in the Lundbeck meta-analyses), the study methodology, including prospective randomisation, patient diagnosis, blinding, study monitoring, and data quality assurance, was designed to provide a valid comparison of the safety and efficacy of the two treatments in the population of interest, in line with EMA/FDA standards.

According to the full scientific assessment report (EPAR),[18] all six trials were submitted when memantine was granted a Community marketing authorisation by the EC. This is when the recommendation was made by the CHMP that memantine's benefits are greater than its risks for the treatment of patients with **moderate to severe** Alzheimer's disease and that memantine be given marketing authorisation for this indication (2005). The recommendation would have been made on the basis of:

1. The Committee's review of all trials' data on quality, safety and efficacy and the fact that all adhered to GCP Guidelines of the Committee for Proprietary Medicinal Products[19] and EU Directive.[20]

2. Acknowledgement that all trials were performed according to the Declaration of Helsinki,[21] and in line with local ethical review board requirements for ethics and informed consent in this special population with dementia.

3. Audits of the trials conducted by relevant Competent authorities.

4. The design of the studies, the qualification of the patients and the selected outcome measures fulfilling the standard requirements of the EMA AD guidelines.[22] (NB: Study MRZ-9605 (Reisberg) was specifically designed and performed in accordance with scientific advice provided by the CHMP).

Furthermore the placebo-controlled, double-blind, parallel-arm study design of the trials MEM-MD-02, MEM-MD-12 and MEM-MD-01 plus MEM-MD-10 was chosen in accordance with the requirements of the FDA Guidelines for the Clinical Evaluation of Antidementia Drugs

and conducted in full compliance with FDA guidelines for Good Clinical Practices and the International Conference on Harmonization (ICH)-Good Clinical Practice and in accordance with the ethical principles that have their origins in the Declaration of Helsinki (which were applicable at the time).

In conclusion, these points provide strong argumentation to support the high standards and quality of data for the clinical trials included in the TAR as well as Lundbeck's meta-analyses. Consequently, it is not justified for the TAR to state that the quality of the trials included for review was *'moderate to poor'.*

Quality of the reporting of the memantine trials

According to PenTAG's own 'Quality Appraisal' of manuscripts (in appendix 3), the publications of only two trials (MD12 and MD01) are taken into consideration when making a generalisation on the reporting quality of memantine trials. The TAR states "*a lack of reporting of key measures of trial quality, thus adding to the uncertainty of the results"* (section 1.5.1, page 35).

It should be noted that all four trials included in the TAR (MRZ-9605, MD02, MD12 and MD01) plus the two additional trials in the Lundbeck meta-analyses (MD10 and 99679) were published in peer-reviewed journals, which contributed to the way in which these trials were reported. Specifically:

1. An evaluation of the impact factors for the manuscript journals indicates that these values are in line with most journals (most journals would have an IF of below 5), with two of the trials included in the TAR (MRZ-9605 and MD02) having a high impact factor. This is indicative of the relative importance of the journals within their own field.

2. The quality of the reporting can be further supported by the fact that the manuscripts for all trials conform to a majority of the recommendations outlined in the CONSORT Statement. Grossberg *et al* (2009)[23] have conducted and published such an evaluation of three of the trials (MRZ-9605, MD01 and MD02) according to the criteria set out in the 2001 CONSORT[24] [25] Statement, further underscoring the fact that these trials indeed conform to the criteria in the majority of the recommendations. A further evaluation of the manuscripts with the current CONSORT guideline (2010)[26] for all six trials confirms the same (Appendix E).

3. The level of detail of data in the manuscripts would have been published in line the specific journal requirements at the time of publication so any missing relevant information is assumed to be attributed to the specific journal restrictions (and not a consequence of poor trial quality).

When reviewing in detail PenTAG's own quality appraisal (in appendix 3 of the TAR) there are aspects of the two publications which are considered *'inadequate'* or *'unknown'* or *'partial'* and yet it is unclear specifically in what way the information is lacking. For example, for point 4 of the quality appraisal, it is difficult to understand why it is considered that the eligibility criteria are *'inadequately specified'* as when looking at the manuscripts and 'participants' column of the data extraction table, there appears to be sufficient detail. For

items which are considered *'inadequate'* or *'unknown'* or *'partial'* by PenTAG's internal review, further detail is provided on these items based on a direct extraction from the clinical study protocols/reports of the MD12 and MD01 trials which, it must be re-iterated, were considered to be of an acceptable standard to the EMA/FDA and other relevant bodies (Appendix F).

The points above are in strong support of the high standards of reporting for the clinical trials included in the TAR as well as Lundbeck's meta-analyses. It is unjustified to conclude that the reporting quality of memantine trials are '*poor*', based on PenTAG's internal quality check of the manuscripts of only two trials (MD12 and MD01). Overall, in addition to the misleading statements made on the quality of the clinical trials, it is further inappropriate for TAR to state that there was a '*lack of reporting of key measures of trial quality, thus adding to the uncertainty of the results'*

4.6.2 Meta-Analysis of Memantine Data

In the submission from Lundbeck the clinical efficacy of memantine within the licensed indication was described using data from the meta-analysis of six trials.[5] The meta-analysis was conducted in accordance with international standards for such analyses (e.g. EMA and Cochrane collaboration). The methods were also in line with the previous MTA conducted in 2004. Within the TAR several points on the meta-analysis methodology were raised.

4.6.2.1 Selection of Included Trials

Within the Lundbeck submission it was stated that the meta-analysis was conducted on pivotal clinical studies. As described previously specific trials included within this analysis were excluded by PenTAG. The TAR included the following comment on the meta-analysis *"Although some details on the methods of analysis were provided, there was no information on how the pivotal trials were ascertained"* (section 4.4.3; page 76).

The inclusion criteria for Winblad 2007 meta-analysis are clearly described within the Lundbeck submission and are in line with key criteria of the PenTAG review: RCT data, a patient population in line with the UK marketing authorization, English language. An additional inclusion criterion for the Winblad 2007 analysis was the availability of endpoint data for four AD domains using standard recognised scales. Full reasons for the exclusion of additional studies are provided in Appendix C of the Lundbeck submission.

Sensitivity analyses were conducted by including three other trials (Asubio IE-2101, Forest MEM-MD-22, and Lundbeck 10112), as although these trials did not meet the complete inclusion criteria for the review they contained some supplementary information on the efficacy of memantine. The results were similar to that of the main analysis.

4.6.2.2 Method of Analysis – LOCF vs. OC

There is a discrepancy between the TAR and Lundbeck submission in regard to the method for analysing data in the trials.

In their review of the data PenTAG state that "*a particular criticism is the use of* [last observation carried forward] *LOCF and* [observed case] *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" (section 1.3.1; page 29). The TAR do not provide a recommendation on what analysis should be used but generally employ the LOCF approach and highlighted a preference for this compared to observed cases (OC). As an example "*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*" (section 4.6.4.3; page 147).

Given the chronic and progressive nature of AD the OC method is considered more appropriate than the LOCF method as LOCF may underestimate or overestimate the treatment effect.[27] The LOCF method can artificially overestimate the clinical state of dropout patients at the end of the study by simulating stability when deterioration is more likely to occur. The under or over estimation of treatment effect is dependent on the balance of withdrawals between the active and control groups. In cases where patients treated with active therapy experience earlier or greater withdrawal than control patients the LOCF method will over estimate the benefit of active drug. In cases where withdrawals are lower or later with active therapy, as is generally the case for memantine, the LOCF analysis will underestimate the effect of drug.[27]

In the primary analyses conducted by Lundbeck an OC approach was utilised. The LOCF analysis was also conducted and this is also presented in the Lundbeck submission (see page 16 to 30 and Appendix D). The conclusions remain unchanged with the use of the LOCF method.

An example of the potential impact of the analysis approach can be seen in the analysis of the trials with memantine on a stable dose of AChEIs on cognition. The TAR utilise an LOCF analysis and for the impact on cognition at 24 weeks conclude "*no overall benefit from combination therapy"* (section 4.8.1.2.4; page 179) (p=0.182). In the analysis conducted by Lundbeck using observed data when the two trials in this population are combined the benefit of memantine approaches statistical significance (p=0.07). The difference between these two results may be due to the approach to the analysis.

4.6.2.3 WMD vs. SMD

Within the TAR data was primarily reported as weighted mean differences (WMD). This is an absolute measure of effect size and in cases where data is being pooled across trials all the data must have been collected using the same scales. The TAR does recognise the SMD approach but states "*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*" (section 4.1.5.1, page 67). This approach restricted some of the analyses that PenTAG could conduct on the data for memantine.

Within the Lundbeck analysis the SMD was used in the meta-analysis allowing assessment of the same outcome even when this has been measured in a variety of ways through the use of different tools. As described above, in accordance with international standards, when combining trials that use different scales to assess the same outcome (as is frequently the case in AD) the SMD is a validated and appropriate method. In fact, given the indication for

memantine extends from moderate to severe AD patients the SMD is an important approach that enables a meta-analysis of data for memantine across the complete licensed indication.

4.6.3 Mixed Treatment Comparison

Within the protocol for the review of clinical effectiveness, it is clearly stated that appropriate comparators are dependent on the severity of the disease (section 3.1.3; page 58). Furthermore, when describing the mixed treatment comparison (MTC) methodology it is stated that "*outputs are presented in terms of treatment effect compared to a common baseline"* (section 4.1.5. 3; page 69).

However, the MTC conducted by PenTAG relied on a pooled dataset for AChEI and memantine regardless of the severity of the included patient populations. For example, MTC of NPI includes data for galantamine (Brodaty 2005) and rivastigmine (Winblad 2007) in mild to moderate patients alongside data for memantine in patients with moderately severe to severe AD (Van Dyck 2007). Data from AChEIs in mild to moderate AD were compared with data from memantine in moderately-severe to severe AD and it is therefore likely that there was a discrepancy in the baseline characteristics of the patients included in these trials, thereby challenging the comparability of the placebo arms between these two populations. Furthermore, no attempts have been made to adjust for this imbalance. This comparison therefore represents a violation of the assessment protocol making the indirect comparison unreliable. The clinical relevance of the comparison is minimal.

4.7 Specific Inaccuracies

Length of Follow-Up

In regard to the length of follow-up of the trials the TAR states: "*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*" (section 1.5.1; page 34).

It should be noted that studies were conducted in accordance with EMA guidelines. The long-term effects of therapy are impossible to determine via prospective placebo-controlled RCTs and there is also the important ethical consideration about continuing to treat patients with high unmet need with placebo once the efficacy of an active treatment has been clearly demonstrated. Longer-term data for memantine is available from observational studies and this was submitted as part of the Lundbeck dossier. However, this evidence was inappropriately excluded from the TAR as described in section 4.4.

Inaccurate Formula for MTC

When describing the pairwise comparisons PenTAG states "*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*" (section 1.1.1.1; page 65). However, the formula provided after the paragraph is incorrect as only the intra-arm variance is accounted for, not the inter-arm (that may occur for instance in case of a dose-response).

Inappropriate Description of Excluded Data Sources

The TAR did not consider many of the data sources included in the Lundbeck review. The TAR makes reference to the discrepancies in the data considered between their review and that conducted by Lundbeck *"The reasons why some studies were included in the Lundbeck analysis but not included in the PenTAG meta-analysis are documented in Appendix 10"* (section 4.4.3; page 76). However, this is misleading as Appendix 10 in the TAR details all of the data considered in the Lundbeck submission and not specifically the trials included in the Winblad 2007 meta-analysis. This thereby suggests that the discrepancy in the included data is greater than is actually the case.

Omission of Key Systematic Review

The search strategy used by PenTAG did not identify at least one systematic review. The Cochrane systematic review for memantine published in 2009 was not included in the TAR.[12] Given the importance of this review in describing the clinical efficacy of memantine this represents a major omission.

Misinterpretation of Scales

In the MTC data is presented for the global deterioration scale (GDS). This includes data for memantine from the Reisberg 2003 study (MRZ-9605). In the GDS a higher score indicates a worse health state. In the study the change from baseline in this outcome was +0.1 for memantine and +0.2 for placebo. The WMD between memantine and placebo is therefore negative (-0.1) indicating a benefit of memantine. In the MTC in the TAR (table 71; page 188) the WMD is negative for memantine but positive for the other therapies (donepezil and rivastigamine). Given the polarity of the GDS scale if would therefore be expected that the clinical benefit of memantine is greater than the other therapies. However, in the MTC the probability of memantine being most effective is lower than the AChEIs. This suggests that the GDS has been misinterpreted.

5 Additional Shortcomings in the PenTAG Approach within this MTA

In addition to the critical concerns raised on the methodology adopted by PenTAG in assessing the clinical effectiveness and cost-effectiveness of AD therapies in England and Wales, we would like to note some further drawbacks of this evaluation.

- Knowledge of the disease was poor throughout the evaluation process
- The appraisal demonstrated a lack of knowledge about the current management of AD patients
- Unpublished evidence, central to the ongoing evaluation of AD therapies, were excluded
- The methods of reviewing industry submissions were not transparent and the reporting of these submissions biased

PenTAG demonstrated a poor application of knowledge on the disease throughout the evaluation process

Although the background section of the TAR reviews evidence on the clinical and economic impact of AD, these considerations were omitted in the clinical and cost-effectiveness sections.

For example, the role of behavioural symptoms on core AD domains, and their impact on healthcare resource use, care-giver burden, rates of institutionalisation and ultimately the cost of AD to healthcare systems and society were presented in the background section. However, PenTAG makes no attempt to distinguish this patient sub-group and reviews all of the evidence by classifying disease severity based on cognitive status. Furthermore, the economic model excludes behavioural symptoms as possible predictors of time to institutionalisation with no scientific rationale provided. Lastly, the evidence submitted by Lundbeck on the efficacy of memantine in the population with APS was omitted from the review in the assessment of both clinical and cost-effectiveness.

PenTAG demonstrated a lack of knowledge about the current management of AD patients

Within their review PenTAG excluded clinical data for memantine in patients on a stable dose of AChEI. This exclusion is inappropriate and reflects a lack of knowledge about how AD patients are currently managed in the UK.
The majority of AD patients have either been treated with AChEIs, are currently treated with AChEIs or are not treated with these therapies as a result of contra-indications and within

the UK the majority of patients with moderate and severe AD will receive AChEI at some point during their disease course. Past or concurrent background therapy with AChEIs does not preclude the prescription of memantine, which is currently the only other agent available for the treatment of AD. Similarly, previous or concurrent use of AD medication is not an exclusion criterion in recent RCTs for novel AD drugs in development.

It is important to note that the EMA chose to evaluate the benefits of memantine based on a pooled evidence base of 6 pivotal RCTs that included patients naive to AChEIs, with a past use of AChEIs or currently treated with a stable dose of AChEIs. By excluding from trials that consider memantine in patients on a stable dose of AChEI the PenTAG takes a contradictory approach to other independent evaluation agencies. Also, given clinical practice in the UK where the majority of patients receive treatment with AChEIs, it is deemed inappropriate to evaluate the cost-effectiveness of memantine in patients who are naïve to any AD medication as this is not a true representation of real-life practice or the target population.

PenTAG adopted an inconsistent and flawed methodology to their appraisal of the health technologies. We have critical concerns on the approach to the evidence review and overall evaluation process in this MTA.

The role of submitted unpublished evidence was unclear. The unpublished analyses that were performed to address the decision problem defined in the scope were omitted from the review. PenTAG acknowledges the importance of the behavioural domain in AD but completely overlooked the evidence submitted by Lundbeck in a sub-group with behavioural disturbances.

- Consultee submissions were reviewed by different people within PenTAG and it is not evident whether a common check-list was used when appraising and reporting on the evidence from the industry submissions. Discrepancies in the summaries across the appraised technologies may have arisen as a result of human error, rather than from major differences in the key quality control points (e.g. quality and relevance of the presented data, adopted methodologies).
- In the TAR, the presentation of the economic models developed by consultees was poor and confusing:
  - There was no acknowledgement of the conservative assumptions and conservative approaches used in consultees' models.
  - No acknowledgement of the strengths of consultees' models was included.
  - The disclosure of evidence by consultees to enable fair recommendations based on the best possible evidence was not recognised.
  - The TAR combined a description of the model with a critique of it, making it impossible for a lay reader to distinguish between the two, or to make a fair judgement on the strengths and limitations of the model.
  - The critique was unclear, inconclusive and poorly presented and lacked the necessary information to enable the consultee to initiate additional analyses in response. The critique leaves the reader feeling that the manufacturers' models are of poor quality when in fact it is PenTAG's description of the model that is poor.

- Most of the criticisms of Lundbeck's model were limited to a discussion of 'missing' details. However this information was included in the Lundbeck submission, either in the main body of the document, the appendices or the references. Lundbeck provided clear citations in all cases.
- There was no discussion of the limitations of the PenTAG model. Importantly, PenTAG did not acknowledge that many of the criticisms levelled at the manufacturer models, applied equally to the PenTAG. The main critique on the model developed by Lundbeck is also applicable to the current PenTAG model.

The methodology employed to appraise the clinical evidence generally lacked reliability:

- The quality of memantine trials was misrepresented. The submitted trials are published in peer-reviewed journals and conformed to the CONSORT guidelines at the time of publication.
- PenTAG provided an inappropriate description of excluded data and did not consider many of the data sources included in the Lundbeck submission.
- The search strategy used by PenTAG did not identify at least one systematic review; a key systematic review for memantine undertaken by the Cochrane Collaboration and published in 2009 was not included.
- PenTAG misinterpreted scales used to assess AD. In the MTC the probability of memantine being most effective is lower than the AChEIs due to the incorrect interpretation of the global deterioration scale, in which a higher score indicates a worse health state.
- PenTAG provided a lack of justification for the employed methodology:
  - While the PenTAG does recognise the SMD approach, it is not appropriately applied to the memantine data thereby restricting the comprehensive review of all available evidence.
  - Despite knowing the limitations of the LOCF in AD[27] PenTAG generally employed this method in their base-case analyses of RCT data.
  - The inclusion of memantine in the MTC analysis is inappropriate due to the differences in the baseline disease severity of patient populations in clinical trials for AChEIs and memantine. The results of such analyses have no clinical relevance. It is meaningless to compare the treatment effect of drugs in mild to moderate AD with those in moderate and severe AD, and inappropriate to use this method as a grounds for selecting the most 'effective' option. Furthermore the analysis violates the assessment protocol, where it is recognised that the appropriate comparators are dependent on the severity of AD.

Appendix A: Comparison of the Lundbeck Model with the SHTAC Model

	Criticism of SHTAC model	Addressed in Lundbeck model	Method used to try and address the criticism	Relevant section of Lundbeck's submission
Alzheimer's disease progression:				
1	Generalisability of risk equations	Yes	Predictive equation re-built using data from the LASER-AD cohort, that was designed to be representative of UK AD patients in terms of gender, severity and residential setting	Appendix O (now published as Rive et al., 2010)
2	Implicit assumption in SHTAC model that FTC = severe Alzheimer's disease	Yes	Severe and FTC are two clearly distinct health states in the model, and patients at the start of the model are all pre-FTC, yet moderate and severe AD are represented.	Section 4.3.2.1, appendix N, page 169
3	Baseline characteristics - change cohort characteristics	Yes	Baseline characteristics of patients form the LASER-AD study	Section 4.3.2.1, appendix N, page 169
Cost data:				
4	Query the costs used: Inaccurate, out-of-date, not UK based	Yes	All resource use included in the model were extracted from the LASER-AD, and unit costs from the most recent PSSRU were applied	

	Criticism of SHTAC model	Addressed in Lundbeck model	Method used to try and address the criticism	Relevant section of Lundbeck's submission
5	pre-FTC too heterogeneous a state for a single cost value	Not relevant	Despite a lower number of patient in pre-FTC compared to FTC in the LASER-AD (moderate to severe AD only), uncertainty parameters (SE) around cost estimates were very similar, showing that from a costing point of view, pre-FTC population heterogeneity was acceptable.	Section 4.3.4.1 Appendix N, page 159-161
6	Query the proportion of people in FTC that are institutionalised	Yes	Using data from the LASER-AD study, 70.4% (69/98) of FTC patients were actually institutionalised. This was used to compute the cost of patient in FTC	Appendix N, page 161
7	Query the exclusion of costs for those in institutionalised care who pay privately	No		
8	No inclusion of carer's costs	No		
Quality of life data:				
9	No daily health benefit associated with treatment	Yes	In order to accurately capture the impact of the quality of life of patients during the pre-FTC state and the developing changes as the patient approaches FTC, utilities in the pre-FTC state were linked to the ADCS-ADL score employing a generalised linear model	Section 4.3.3, appendix N, page 171
10	No benefit for those going straight from pre-FTC to death (related to above point)	Yes	As above	Section 4.3.3, appendix N, page 171

	Criticism of SHTAC model	Addressed in Lundbeck model	Method used to try and address the criticism	Relevant section of Lundbeck's submission
11	pre-FTC too heterogeneous a state for a single utility value	Yes	As above	Section 4.3.3, appendix N, page 171
12	Query the values used	Yes	Utility weights were estimated on the UK sample of the LASER-AD cohort, using a mapping of the EQ-5D and applying UK tariffs. Comparison with previously published estimates showed high consistency across sources.	Section 4.3.3, appendix N, page 131-163 Section 4.6
13	No inclusion of carer's quality of life	No	No relevant data source identified	
Treatment and effectiveness:				
14	Assume treatment stops once enter FTC	Not relevant	This criticism was related to consistency between treatment stopping rule and AChEIs indication (mild to moderate, not severe). This therefore does not apply to memantine whose indication is form moderate to severe AD. Memantine was assumed to be administered as long as patients remained in the pre-FTC state.	Section 4.2.6
15	No consideration of treatment drop-out, non-responders, adverse events	Not relevant	The economic model does not include considerations of dropouts from treatment, because the drop out rates were similar between memantine and placebo (while drop out tends to be higher for AChEIs). In order to have a proxy of the impact of drop out, sensitivity analyses were performed using both LOCF or OC methods and did not show major	Section 4.6

	Criticism of SHTAC model	Addressed in Lundbeck model	Method used to try and address the criticism	Relevant section of Lundbeck's submission
			differences between the two approaches thus confirming that the non inclusion of drop out in the memantine model does not lead to an overestimation of the benefit	
16	No treatment effect observed in psychiatric symptoms	Yes	The new predictive equation includes the standard and validated psychiatric symptoms assessment scale NPI as a predictor of time to FTC. Treatment effect measured with this scale is then incorporated and translated into reduction in risk of reaching FTC using the predictive equation.	Appendix O (now published as Rive et al., 2010), section 4.3.2.1, appendix N page 168-170
17	No treatment benefit beyond 6 months	Not relevant	Treatment effect was extracted for 6-month clinical trials and assumed to be sustained over time. Since other data may support longer effect of Memantine the model may be conservative in that regard.	Section 4.6, Atri et al., 2008 Ferris et al., 2001
18	Placebo effect observed in trials	No		
19	Responder analyses not included	No	Treatment effect computed on the overall population, regardless of whether patients were responders or non-responders	
Modelling:				
20	Time horizon longer than 5 years	Not	Time horizon was 5 years. However, it should be noted that after 4 years, the proportion of pre-FTC patients fell below	Markov cohorts in appendix N,

	Criticism of SHTAC model	Addressed in Lundbeck model	Method used to try and address the criticism	Relevant section of Lundbeck's submission
		relevant	0.01% for both treatment strategies (memantine and usual care. Given that no difference in mortality was assumed between treatments, no change on total incremental cost or effectiveness will occur beyond 5 years, meaning that time horizon was sufficient to capture all long-term costs and effectiveness of the drug.	page 179-180 Death probability section 4.3.2.2
21	Constant mortality assumed	Yes	Monthly death probability was derived from pre-FTC patients in the LASER-AD cohort. Because of increasing death rate with time, a Weibull parameterisation was chosen. Mortality assumed to be the same for treated and untreated patients.	Section 4.3.2.2, appendix N page 170
22	Over-estimated' mortality	Yes	Mortality estimated from the LASER-AD (representative of AD patients in the UK), restricted on the modelled population (pre-FTC patients)	Section 4.3.2.2, appendix N page 170
23	Lots of queries regarding the PSA	Yes	Distributions around baseline characteristics, treatment effects, utility in the FTC health state, and the costs per health state were used. All distribution parameters derived from observed data, no assumption was required.	Section 4.4.1, appendix N page 172-173
24	Inclusion of multi-way sensitivity analyses	To an extent	Only sensitivity analysis involving multiple parameters was the probabilistic sensitivity analysis (see above)	Section 4.4.1, appendix N page 172-173
25	Individual vs population characteristics	To an	Specific population of patients with Agitation/Aggression and/or Psychotic Symptoms (APS) assessed using same	Section 4.5.2.2, appendix N

	Criticism of SHTAC model	Addressed in Lundbeck model	Method used to try and address the criticism	Relevant section of Lundbeck's submission
		extent	model, with baseline characteristics and treatment effect specific to this population	pages 168-170
26	No monitoring of MMSE/ADL etc – cannot model current NICE guidance	Not relevant	This criticism was related to AChEIs indication (mild to moderate, not severe). This therefore does not apply to memantine whose indication is form moderate to severe AD. However, it should be noted that current model allows such monitoring (this is for instance used to estimate utility in pre-FTC that evolves with ADCS-ADL score)	Section 4.3.3, appendix N, page 171
27	Accounted costs during initial treatment period, but not any health benefits	Yes	Both treatment effect and associated costs accounted for since the beginning of the model, until patients reach the FTC state	Section 4.3.2.1, Section 4.3.4.2, Section 4.2.6

Appendix B: Comparison of the Lundbeck Model with the PenTAG Model

The following table provides a summary of the key critiques of the Lundbeck model raised by PenTAG. The table lists these items and considers these alongside the assumptions from the PenTAG model.

Table 1: Comparison of the Lundbeck and TAR Model

	Critique on Lundbeck's model	Assumption in PenTAG model	Comments
Constructing equation of time to 'event" Lundbeck model – event was defined as FTC PenTAG model - event was defined as institutionalisation	**PenTAG statement, p 214**: "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." **Lundbeck**: When developing the predictive equation of time to FTC in AD patients, univariate analyses were run to identify the potential predictors. Neither AChEIs nor antipsychotics were significant and therefore were not included in the multivariate analyses or the final model. For detailed description of the construction of the predictive equation, please see Appendix O, p 315-317. One could argue that effect of AChEI could have been deducted from the baseline characteristics of patient population at the model start (in a similar fashion as memantine effect was modelled). This would be in favour of memantine. We avoided such	**PenTAG statement, p 283** "Although MMSE and Barthel-ADL were not identified as statistically significant variables in explaining the variance of time to end of pre-institutionalisation, both were retained in the model so that a treatment effect could be incorporated into the decision model." **Lundbeck:** This lack of statistical significance for the main predictors of the time to institutionalisation (MMSE and Barthel index) may raise several issues, the main one being that this implies a large variability in the estimation of coefficients associated to these parameters. A direct consequence of this is that there is a high uncertainty regarding the "true" value of these coefficients, which may result in inaccurate predictions.	Unjustified critique on Lundbeck's model, please remove the following statement from the TAR **p 214:** "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."

	Critique on Lundbeck's model	Assumption in PenTAG model	Comments
	an exercise as it was deemed to be inappropriate for the target population in question for whom the relevant comparator is the best supportive care with or without background AD therapy.		
Constructing equation of time to 'event" Lundbeck model – event was defined as FTC PenTAG model - event was defined as institutionalisation	**PenTAG statement, p 215:** "The predictive equation has not be validated against an external data source, therefore the degree to which the results are generalisable is unclear." **Lundbeck:** The predictive equation has not been validated against an external data source due to absence of such source. Please see comments on validation of the equation in section 2 of the response.	**Lundbeck:** The equation developed by PenTAG has not been validated against an external source either. No description of this exercise was provided in the TAR. Furthermore we have concerns with the statistical approach adopted by PenTAG, also see Section 2.1	Unjustified critique on Lundbeck's model, please remove the following statement from the TAR: **p 215:** "The predictive equation has not be validated against an external data source, therefore the degree to which the results are generalisable is unclear."

	Critique on Lundbeck's model	Assumption in PenTAG model	Comments
Modelled cohort	**PenTAG statement, p 214:** "It is unclear how representative the patient sample is with respects to the general moderate to severe AD population in the UK." **Lundbeck:** The LASER-AD cohort comprised an epidemiologically representative sample of patients, who were treated in accordance to the established clinical practice in the UK. e.g. Ryu et al., 2005[28]. The PDF was supplied along with the submission documents	**PenTAG statement, p 266:** "Note that the data informing disease progression are that from a prevalent cohort of patients living in the community, and is therefore not fully representative of the target population of patients in England and Wales living in the community and in institutionalised 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 institutionalised was available from the LASER-AD study. Livingston and colleagues"	Unjustified critique on Lundbeck's model, please remove the following statement from the TAR: **p.214:** "It is unclear how representative the patient sample is with respects to the general moderate to severe AD population in the UK."
Programming and reporting	**PenTAG statement, p 215: "** The programming of the statistical model is poorly described, meaning there is concern that it may not have been used appropriately." **Lundbeck:** The TreeAge model and the user guide have been submitted. This simple 3-state model has been programmed in the standard software. We will be glad to examine and correct (if relevant) any identified inconsistencies, if any, otherwise, we insist on removing this unjustified statement from the TAR. Also, the quality of programming was controlled using a double independent programming procedure	**Lundbeck:** PenTAG did not provide the user-guide for the model. The model was poorly described in the TAR. An in-depth review of the model was undertaken to understand many modelling assumptions that were not document in TAR. For instance, it was not possible without the electronic model to understand that PenTAG actually made the assumption that ADCS-ADL_{19} and ADCS-ADL_{23} were identical when converting memantine benefit into Barthel index (resulting in an under-estimation of memantine effect, see section 2.1.3.3 "mapping of functional scales").. Also, a different precision of input parameters were	Unjustified critique on Lundbeck's model, please remove the following statement from the TAR: **p 215:** "The programming of the statistical model is poorly described, meaning there is concern that it may not have been used appropriately."

	Critique on Lundbeck's model	Assumption in PenTAG model	Comments
	(see page 33 of Lundbeck dossier: "To assure quality control, the model was independently verified by a second independent modeller reconstructing the analysis based on the formal assumptions and inputs of the model. Results obtained were strictly identical in non-stochastic analyses and similar in stochastic analyses although programming was different.")	used in the electronic model (intercept terms from lambda coefficients of predictive equations, Barthel baseline score), preventing a rigorous re-building of the model.	
Mapping between the scales	**PenTAG statement, p 216:** "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." **Lundbeck:** All relevant information on transformation algorithm was provided Appendix N, p 163. SIB scores were transformed into ADAS-cog scores using a linear regression model computed on data from the LASER-AD study data. Consequently, this led to the building of the following equation between the two scores was constructed: ADAS-cog = 83.0831 - 0.5745 * SIB. R^2 of the model was 75.6%, thereby indicating good predictive properties.	**PenTAG p 276:** "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." **Lundbeck:** Insufficient information on the mapping procedures is provided to allow replication of the exercise. In particular: - The exact correspondence between Barthel scale items and ADCS-ADL_{19} items (which item from ADCS-ADL_{19} was used as a proxy for which item from Barthel scale, and which items from Barthel scale had no single equivalent in ADCS-ADL_{19}) is not documented. - The exact correspondence between ADCS-	Unjustified critique on Lundbeck's model, please remove the following statement from the TAR: **p 216:** "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."

	Critique on Lundbeck's model	Assumption in PenTAG model	Comments
		ADL_{23} items and ADCS-ADL_{19} items is not documented. - The "which most closely correlated" criterion used to match items from two different scales is not explicitly described. The "acknowledgement" section of the TAR tends to indicate this is more related to expert opinion than statistical correlation techniques, but no confirmation of this can be found in the report. Please see Section 2.1 for further comments	
Mapping between the scales	**PenTAG statement, p 217:** "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." **Lundbeck:** rescaling scores of one of the scales (ADCS-ADL_{19}) into scores of the other one (ADCS-ADL_{23}) simply assumes that: - When the minimum score is achieved on one scale it is also reached on the other	**Lundbeck:** PenTAG did similar mapping exercise to convert ADCS-ADL_{19} and ADCS-ADL_{23} into Barthel Index and makes the same assumption when when mapping ADCS-ADL_{19} and Barthel (TAR, page 277: "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."). Implication of the mapping on results were not tested or discussed in the TAR.	Unjustified critique on Lundbeck's model, please remove the following statement from the TAR: **p 217:** "The Reisberg and van Dyck study measured functional status using the ADCS-ADL19 (scores ranging between 0-54), not the ADCS-ADL23 (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."

	Critique on Lundbeck's model	Assumption in PenTAG model	Comments
	- When the maximum score is achieved on one scale it is also reached on the other - Between the two, the relation is linear Regarding the first two points, it should be noted that the exact same assumptions were chosen by PenTAG when mapping ADCS-ADL_{19} and Barthel (TAR, page 277: "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."). The last point can be justified by the very similar content of the two scales, in particular the inclusion of both basic and instrumental ADLs that allow detecting functional disability in the early (instrumental) and late (basic) stages of the disease.		
Modelling treatment effect	**PenTAG statement, p 216:** "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 <u>were meta-analysed and literally added to the related baseline variables in the risk equation</u>".	**PenTAG statement, Appendix 16, page 151:** "It is then assumed that decline after time t continues at the same rate as that in the untreated individual, but that the treated individual is constantly x points above the untreated individual (see explanation of treatment effect for the one dimensional Markov model below for discussion of this assumption if the Mendiondo and colleagues54 disease progression eqn is used)." **Lundbeck:** The exact same assumption was made	Unjustified critique on Lundbeck's model, please remove the following statement from the TAR: **p 216:** "Specifically, changes on the ADAS-cog baseline, ADCS-ADL baseline and NPI baseline scores <u>were meta-analysed and literally added to the related baseline variables in the risk equation</u>".

	Critique on Lundbeck's model	Assumption in PenTAG model	Comments
		in Lundbeck model evaluating the cost-effectiveness of memantine, by considering an added benefit on the different symptoms, but no effect on the rate of decline (slopes in the predictive equation).	
Mortality	**PenTAG statement, p 218: "**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." **Lundbeck:** In an attempt to answer PenTAG concern regarding the death probability, an additional sensitivity analysis was conducted by doubling the death probability in the model. As a result, proportion of dead patients at the end of the 5-year evaluation period for both treatment strategies increased to 74.5 compared to 49.8% in base case analysis. This resulted in lower costs,	**PenTAG, p 271** *"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 institutionalisation, but has no impact on survival."* **Lundbeck:** The exact same assumption was applied in Lundbeck memantine model, based on the same rationale (avoid a treatment effect on survival that has never been demonstrated). In both the Lundbeck and the PenTAG models, mortality rate was computed using cohorts of patients in the target population (pre-FTC and living in community respectively) and was assumed to remain the same when patient's health state changed (transition to FTC or institutionalisation respectively) to avoid creating an artificial benefit of treatments (through delayed progression to FTC or institution respectively) that has never been demonstrated. It should also be noted that in both data source used, patients were still followed when	Unjustified critique on Lundbeck's model, please remove the following statement from the TAR: **p 218: "**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."

	Critique on Lundbeck's model	Assumption in PenTAG model	Comments
	QALYs and time in pre-FTC for both treatment arms. However, memantine was still associated with 0.029 additional QALYs compared to standard care (0.031 in base case analysis), 4.9 additional weeks in pre-FTC (5.6 weeks in base case) and cost savings of £1,360 (£1,711 in base case analysis). In corresponding stochastic analysis, probability of memantine being more effective was 99.7% (99.8% in base case analysis) and probability of memantine being less costly was 94.7% (96.4% in base case analysis). These results therefore fail to fully corroborate PenTAG conclusion regarding influence of mortality on results, indicating instead that, even if outcomes per treatment strategy are indeed sensitive to mortality rate, conclusions of the model in terms of difference between treatments remains stable, even when drastically (doubling) increasing the mortality rate.	reaching the next health state (pre-FTC or institutionalisation respectively), so that mortality estimates also incorporates potential increased mortality associated with this changes. In the light of these two identical approaches, the reason of PenTAG critique is unclear.	
Collection of RU	**PenTAG statement, p 218. "**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	**PenTAG statement, p 266: "**The 1997/8 UK-based study by Wolstenholme and colleagues181 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	Unjustified critique on Lundbeck's model, please remove the following statement from the TAR: **p 218. "**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

	Critique on Lundbeck's model	Assumption in PenTAG model	Comments
	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." **Lundbeck:** The LASER-AD study was published in various publications; most relevant were submitted to NICE. For more information please see: Livingston G, Katona C, Roch B, et al. A dependency model for patients with Alzheimer's disease: its validation and relationship to the costs of care-the LASER-AD Study. Curr Med Res Opin 2004, 20:1007-16	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." **Lundbeck:** No further details are provided on how data was collected, recall period, handling missing data, etc	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"
Drug cost	**PenTAG statement, p 218.** "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."	**PenTAG statement, p 299:** Monthly drug costs were calculated from costs reported in the BNF 58 for the specific doses of interest." AND "Note that the relevant drug costs do not differ between BNF58 (4th quarter 2009) and BNF59 (1st quarter 2010)."	Unjustified critique on Lundbeck's model, please remove the following statement from the TAR: **p 218.** "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."

	Critique on Lundbeck's model	Assumption in PenTAG model	Comments
	Lundbeck: The relevant of this critique is questionable. The unit cost of Memantine was taken from BNF 58 (September 2009), the latest available issue at the moment of submission. (BNF59 was issued in March 2010). Differences can be explained by the fact that Lundbeck considered the price at the ex-manufacturer level (i.e. sold to the wholesaler) while PenTAG considered the price sold to pharmacists. Furthermore TAG indicates that "this increased cost had little bearing on the base-case cost-effectiveness results". However the statement that "clearly its importance will be magnified if other changes are simultaneously made to the model, such as lessening the effect of Memantine" is ambiguous and not supported by evidence.		
Utilities	**PenTAG statement, p 220: "**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." **Lundbeck:** This was done to address the limitation of SHTAC 2004 model by allowing gradual decrease of utility for patients in pre-FTC, in order to be able to also capture the benefit of the treatment in patients who did not survive long enough to reach FTC state. The same was done in the current	**PenTAG statement, p 263:** "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."	Unjustified critique on Lundbeck's model, please remove the following statement from the TAR: **p 220: "**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."

	Critique on Lundbeck's model	Assumption in PenTAG model	Comments
	PenTAG model		

PenTAG statement, p 221: "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
- to transform some outcome scores from one scale (from the RCTs) to other scales (which were specified in the risk equation)."

We believe that this table clearly shows that the above critique of the Lundbeck's model is unjustifiable on any of the above statements. We request to remove any unjustified statements from the TAR.

Appendix C: Responses to Specific Criticisms of the Lundbeck Memantine Model

This section presents the responses to the PenTAG comments on Lundbeck's model to assess the cost-effectiveness of memantine in moderate and severe AD. The responses are organised below in the order of their discussion in the TAR.

The Decision Problem

MAJOR COMMENTS

PenTAG statement "*The patient cohort consists of individuals with moderate to severe AD as measured using a number of functional and behavioural instruments, but not MMSE"* and *"All individuals are assumed to have moderate AD, as defined in Table 80."*

These are incorrect statements. The following wording should be used instead:

- "The baseline patient population were moderate or severe AD patients with MMSE score of 19 and below, who did not yet require or receive full-time care, i.e. independent patients living in the community"[6].

From the Lundbeck submission (Appendix N , page 168): "The values for the baseline parameters and slopes were derived from the LASER-AD cohort, restricted to moderate to severe (MMSE <20) patients who were neither dependent nor institutionalised (ie in pre-FTC) at baseline"

PenTAG statement, page 216"*The manufacturer of memantine submitted a model-based economic evaluation comparing it with no pharmacological treatment*."

This is an incorrect statement. As stated on page 31 of the dossier submitted by Lundbeck: "The comparator for memantine was standard care, which has been defined as any treatment received for AD. In the UK, for moderate patients, this could be AChEIs or no therapy and for severe patients, this would be no therapy. Standard care for patients not on an AChEI is considered to be receiving social support and assistance with day-to-day activities."

[6] A need for FTC was a multidimensional endpoint in the model and combined an assessment of patient's dependency status and location of care. The former was determined based on assessment of psychical and functional disability as measured by the Cooperative Study – Activities of Daily Living Scale (ADCS-ADL), applying the qualitative classification by Livingston et al 2004.

The following wording is proposed: **"**The manufacturer of memantine submitted a model-based economic evaluation comparing two treatment alternatives, memantine and standard care, defined as no treatment or any background AD therapy".

PenTAG statement, page 211 "*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). In the current submission This point is acknowledged in Lundbeck's current submission."*

This is an ambiguous statement. The following wording is suggested:

"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). In the current submission the definition of the sub-population was refined on the ground of clinical expertise (appendix B Consensus Statement on APS Sub-group Definition)"

MINOR COMMENTS

PenTAG statement, page 210 "*The model is based on a Markov approach and health outcomes were expressed as QALYs.*"

The following wording is proposed: "The model is based on a Markov approach and health outcomes were expressed as time to FTC and QALYs."

An Overview of How the Model Works

MAJOR COMMENTS

PenTAG statement, page 211 "*All individuals are assumed to have moderate AD, as defined in Table 80.*"

This is an incorrect reference. Table 80 presents the baseline characteristics of the memantine cohort. The table should present the baseline characteristics of the underlying population, as observed in LASER-AD study, i.e. 'standard care arm'. This is provided in Appendix N, page 168 of the Lundbeck submission and is replicated below:

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
Symptomatic 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

PenTAG statement, page 211 "*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.*"

It is an ambiguous statement that leaves the reader with no conclusion.

The following is taken from the Lundbeck dossier: "The need for FTC is a relevant outcome in moderate to severe AD. Patients requiring FTC present a major burden to carers and the healthcare system as AD is associated with significantly lower cognitive and functional abilities."

The definition of FTC was becoming dependent or institutionalised. The results of the LASER-AD study supported the relevance of loss of independence as a main driver of cost and utility[28] [29] (and provided information regarding institutionalisation of patients, as well as validating the classification of dependency based on functional performance. This is deemed to be a more relevant approach than predicting time to institutionalisation alone.

MINOR COMMENTS

PenTAG statement, page 211 *"The model is run probabilistically, although not all of the appropriate variables are specified as distributions."*

As stated in the dossier submitted by Lundbeck (page 36): "For the probabilistic analysis, normal distributions around baseline characteristics, treatment effects, utility in the FTC health state, and gamma distributions around the costs per health state were used (distribution parameters are reported in the tables above)."

No suggestion is provided by PenTAG regarding the potential additional appropriate variables for which distributions should have been implemented. This conclusion therefore appears unjustified.

The following wording is suggested:

"The model is run probabilistically, by associating a priori distributions to baseline characteristics, treatment effects, utility in the FTC health state, and costs per health state."

PenTAG statement, page 212 "*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*."

The statement above is unclear. Clarification is required to answer to this comment.

As stated in the dossier, stochastic analyses led to the conclusion that memantine was less costly (i.e. cost-effective at a willingness-to-pay of £0 per QALY) compared to standard care in 96.4% of Monte-Carlo simulations and more effective (i.e. cost-effective for an infinite willingness-to-pay) in 99.8% of simulations (page 38 of submitted dossier). Between these two values, the probability of memantine being cost-effective compared to standard care gradually increases with the willingness-to-pay.

PenTAG statement, page 212 "*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*."

The statement above is unclear.

The following wording is suggested "Of the 224 patients included in the LASER-AD study, 33 (14.7%) were dependent at baseline, 23 (10.3%) were institutionalised and 51 (22.8%) were both dependent and institutionalised. 117 remaining patients were analysed to derive the transition probabilities of going from pre-FTC to FTC in this model".

Comparator Treatment Options

MAJOR COMMENTS

PenTAG statement, page 212 "*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*."

As described in Sections 2.6 "Place of Memantine in the Treatment Programme for AD in the UK" the submission from Lundbeck presented evidence to support a restricted recommendation for memantine in patients in whom AChEIs is not deemed to be an optimal treatment strategy, that is:

- Moderate AD patients withdrawn from AChEIs;
- Moderate AD patients contraindicated for AChEIs;
- Moderate patients requiring adjunct treatment while on stable dose with AChEIs; and
- Patients with severe AD.

Therefore in the presented economic evaluation, memantine was compared with standard care for these patients in the UK setting, i.e. best supportive care with or without background AD therapy. The cohort-level Markov simulations were based on the data from the LASER-AD study. The LASER-AD cohort is an epidemiologically representative sample of patients treated in accordance to the established clinical practice in the UK (Ryu et al., 2005[14] - the full reference was supplied along with the submission documents).

Lastly, the submission presented the data for a sub-group of patients with APS. Treatment with AChEIs would not be a relevant comparator for this group of patients either.

The following wording is suggested "The comparator for memantine was standard care, which has been defined as any treatment received for AD. In the UK, for moderate patients, this could be AChEIs or no therapy and for severe patients, this would be no therapy. Standard care for patients not on an AChEI is considered to be social support and assistance with day-to-day activities."

The Risk Equation

MAJOR COMMENTS

PenTAG statement, page 214 "*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*."

When developing the predictive equation of time to FTC in AD patients univariate analyses were run to identify potential predictors. Neither AChEIs nor antipsychotic treatments were significant predictors and therefore were not included in the multivariate analyses of the final model. For a detailed description of the construction of the predictive equation please see Appendix O, page 315-317 of the submission documents.

One could argue that the effect of AChEI used could have been deducted from the baseline characteristics of the patient population at the model start (in a similar fashion as the memantine effect was modelled). This would generate a more favourable outcome for memantine. Lundbeck avoided such an exercise as it was deemed to be inappropriate for the target population in question for whom the relevant comparator is best supportive care with or without background AD therapy.

PenTAG statement, page 214 "*It is unclear how representative the patient sample is with respects to the general moderate to severe AD population in the UK*."

The LASER-AD cohort comprised an epidemiologically representative sample of patients, who were treated in accordance to the established clinical practice in the UK. e.g. Ryu et al., 2005[14] - the full reference was supplied along with the submission documents

PenTAG statement, page 214 "*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*."

The model relies on a validated functional classification of patient's dependency, which combines both basic and instrumental activities of daily living. The model was first developed in a Belgian cohort and then independently validated for the UK sample. For more details please see the publication by Livingston et al., 2004[28] (the full reference was supplied along with the submission documents). In brief, the model does not rely on a specific threshold, but rather on the automatic classification algorithm. This is a validated model. Therefore proposed sensitivity analyses on alternative assumptions regarding the definition of dependence, i.e. threshold or classification systems, are not deemed to be relevant.

Also dependency was found to be a major driver of both cost[28] and utilities[29] even when accounting for residential setting (community or institution), while severity (cognition) was not.

PenTAG statement, page 215 "*The predictive equation has not be validated against an external data source, therefore the degree to which the results are generalisable is unclear*"

The predictive equation has not been validated against an external data source due to the absence of such source. Importantly, the same limitation applies to the PenTAG model which has also not been validated against an external source.

It should be noted, however, that the model shares the main predictors with the original model by Stern et al[30]and its successor.[2] Similarly to these models, the current equation takes into account patient's cognition disability and behavioural symptoms, yet it also includes a functional domain, which determines patient dependency status and thus a need for FTC.[28] The present analysis shows that baseline cognitive impairment, functional disability and behaviour disturbances are found to be the main static predictors of time to FTC.

The predictive equation employed in the Lundbeck model addresses some of the limitations of the original model by Stern et al [30]and its successor.**Error! Bookmark not defined.**[2] It is more consistent with the clinical course of AD and allows for the inclusion of evidence on all measurable clinical manifestations of AD, i.e. cognition, functioning and behavioural, using standard validated scales to assess each domain.

Notably, the previous models found behavioural symptoms to be a predictor of time to FTC despite being based on a patient cohort with predominantly mild AD. This may suggest the underlying patient population had rapidly declining AD.[31] [32] The model developed here is based on an epidemiologically representative sample of patients in order to widen its application to other settings. It identifies the speed of patient deterioration through dynamic predictors and, by extension, reflects differences in disease progression. The analysis shows that speed of cognitive decline and functional impairment are independent predictors of time to FTC. This feature of the equation allows the evaluation of both symptomatic treatments and also potential disease-modifying effects. Importantly, a linear decline in disease progression factors is assumed, which may not necessarily reflect the natural history of the condition, yet fits with current recommendations.[33] [34] The model shows excellent goodness of fit confirms the validity of the chosen mathematical functions in the equation.

Finally, unobserved heterogeneity was assessed in Lundbeck's predictive equation to identify whether some unobserved factors could affect progression to FTC in some individuals (see appendix O, and Rive et al., 2010). The addition of an unobserved heterogeneity component to the final predictive equation had negligible impact on the estimation of model coefficients: there was a 2.4% change in the estimation of NPI coefficient and <1% change in other coefficients. Heterogeneity was insignificant (p=0.967).

PenTAG statement, page 215 "*The programming of the statistical model is poorly described, meaning there is concern that it may not have been used appropriately*"

The TreeAge model and the user guide were part of the Lundbeck submission. The simple 3-state model was programmed in the standard software. Lundbeck will be very happy to

examine and correct (if relevant) any identified inconsistencies. If no such inconsistencies have been identified we strongly request that this unjustified statement is removed from the TAR.

PenTAG statement, page 215 "*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*"

Slopes in the predictive equation are monthly rates of change of the assessment scales (i.e. the difference between the score at a given visit and score at baseline divided by the number of months elapsed since baseline) (Please see appendix O, page 315, or Rive et al., 2010 for more details). In the construction of the predictive equation these were allowed to be time-dependent variables. However, they were considered constant over time for the purpose of the model programming i.e. assuming linear decline in keeping with regulatory and expert recommendations.[33] [34] For the computation of the baseline characteristics of the standard care cohort, these slopes of decline were estimated using a repeated measurement regression analysis with time as the fixed effect and patients as the random effect. Estimates (e.g. -0.7503 for ADCS-ADL_{23}) then represent the average monthly decline of moderate to severe pre-FTC patients on this scale. This last component was indeed not specified in the submission.

PenTAG statement, page 215 "*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*"

Please see the response provided above**.**

Noteworthy, the approach suggested by PenTAG is methodologically unsound. Slopes of continuous linear decline on ADCS-ADL and ADAS-cog are included in the predictive equation (all details in Appendix O, or Rive et al., 2010) and reflect the degenerative nature of the disease. Following the approach suggested by PenTAG would then require updating predictors in the equation at each time cycle because patient health state continuously deteriorates. This would imply using only the shortest-term part of the equation, while the predictive equation developed for the Lundbeck model was based on survival modelling

methods in order to ensure the best possible accuracy of predicted survival on the overall time period of observation (4.5 years). Please section 2.1 in this response.

Estimating Relative Treatment Effects

MAJOR COMMENTS

PenTAG statement, p 216 "*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*".

The same assumption was employed in the SHTAC model in 2004 and in the current PenTAG model, as stated in Appendix 16, page 151 of the TAR "*It is then assumed that decline after time t continues at the same rate as that in the untreated individual, but that the treated individual is constantly x points above the untreated individual (see explanation of treatment effect for the one dimensional Markov model below for discussion of this assumption if the Mendiondo and colleagues disease progression eqn is used).*"

PenTAG statement, p 216. "*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*."

As stated in the submission on page 18, mild patients were excluded from all analyses on the basis of the MMSE score at baseline. Only patients meeting the criterion of the current marketing authorization (MMSE<20) were included.

The following wording is suggested "In three of the studies, patients were said to have mild to moderate AD. The submission acknowledges this and states that mild patients (MMSE 20 or above) were excluded from the analysis."

PenTAG statement, p 216. "*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"*.

This is an incorrect statement. Given the chronic and progressive nature of AD the OC method is considered more appropriate than the LOCF method as LOCF may underestimate or overestimate the treatment effect.[27] The LOCF method can artificially overestimate the clinical state of dropout patients at the end of the study by simulating stability when deterioration is more likely to occur. The under or over estimation of treatment effect is dependent on the balance of withdrawals between the active and control groups. In cases where patients treated with active therapy experience earlier or greater withdrawal than control patients the LOCF method will over estimate the benefit of active drug. In cases where withdrawals are lower or later with active therapy, as is generally the case for memantine, the LOCF analysis will underestimate the effect of drug.[27]

Furthermore, the sensitivity of the model to these input parameters has been tested, please see Appendix N, page 187 of the Lundbeck submission. Employing LOCF values yielded only

minor differences with the base-case model and did not change the overall model conclusions.

The following wording is suggested:

"*The meta-analyses used to estimate base-case treatment effects were all based on observed case analysis. Alternatively, LOCF approach was used in sensitivity analyses and revealed no impact of imputation method on model outcomes*."

PenTAG statement, p 216. "*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*."

This is an inappropriate statement. The above trials were conducted in patients with moderately severe to severe AD and thus are not entirely in accordance with the decision problem on stated page 30 of the Lundbeck submission: "The model estimates the cost-effectiveness of memantine plus usual care versus usual care alone in the UK for patients with moderate to severe AD, as well as for the APS sub-group, as this sub-group has higher medical needs and incurs higher use of resources."

As described in section 4 the remit of the TAR was to review the effectiveness of the AD therapies within their licensed indications. For memantine this is moderate to severe patients who are naive to AD medication, or have a history of past use of AChEIs or who are currently on a stable dose of AChEIs. A meta-analysis of six large multicentre RCTs, which was conducted in accordance with internationally recognised standards (e.g. EMA and Cochrane) and has been widely published in the peer-reviewed literature, provides the evidence for the clinical efficacy of memantine in this licensed population. In fact, in 2005, the EMA extended the indication for memantine to moderate and severe AD on the basis of this meta-analysis. The approach to this meta-analysis in terms of pooling trials based on concomitant AChEI therapy was in line with the previous TAR for AD conducted in 2004. Despite the availability of this evidence the TAR did not consider the meta-analysis in their review and assessed only two of the six included trials. The TAR review of memantine therefore does not cover the full licensed indication, although importantly this is not explicitly stated in the report.

It is important to note that Lundbeck did not state decision problem: as memantine monotherapy compared with placebo alone.

PenTAG statement, p 216. "*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*."

All relevant information on the transformation algorithm was provided in Appendix N, page 163 of the Lundbeck submission. SIB scores were transformed into ADAS-cog scores using a linear regression model computed on data from the LASER-AD study data. Consequently, the following equation between the two scores was constructed:

ADAS-cog = 83.0831 - 0.5745 * SIB.

R^2 of the model was 75.6%, thereby indicating good predictive properties.

PenTAG statement, p 216. "*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.*"

This is an unclear statement. Clarification will be needed in order to understand the analyses performed by TAG and their results. It is not possible to deduce this from the statement above.

PenTAG statement, p 217. "*The Reisberg and van Dyck study measured functional status using the ADCS-ADL19 (scores ranging between 0-54), not the ADCS-ADL23 (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.*"

The rescaling of scores from one of the scales (ADCS-ADL_{19}) into scores of the other scale (ADCS-ADL_{23}) simply assumes that:

- When the minimum score is achieved on one scale it is also reached on the other
- When the maximum score is achieved on one scale it is also reached on the other
- Between the two, the relation is linear

Regarding the first two points, it should be noted that the exact same assumptions were chosen by PenTAG when mapping the ADCS-ADL and Barthel scores (see page 277: "*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.*").

The last point can be justified by the very similar content of the two scales, in particular the inclusion of both basic and instrumental ADLs that allow for detection of functional disability in the early (instrumental) and late (basic) stages of the disease.

PenTAG statement, p 217. "*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."*

Setting the treatment effect on functional domain to 0 is unlikely to test the robustness of the model assumptions around the rescaling algorithm for ADCS-ADL_{19}. The submission presents the set of sensitivity analyses (Appendix N, page 193), where memantine treatment effect was varied within plausible ranges, (i.e. lower and upper limits of confidence intervals around the mean estimates). None of these analyses yielded estimates

largely different from the base-case. The model conclusions remained unchanged in all analyses.

Furthermore, when we rerun the model using the estimate proposed by TAR (memantine treatment effect on functioning as measured by ADCS-ADL_{19} in a restricted meta-analysis including only two out of the 6 available studies; WMD 1.408, $p<0.05$), despite the unjustified selection of the studies proposed by TAR, the model remained stable producing similar results to that from the base-case. None of the model conclusions changed.

The above analyses prove the robustness of the Lundbeck model to the changes of treatment effect on functioning and assumptions around rescaling algorithm.

Importantly, the statement "Meaning it is not all clear memantine monotherapy is associated with improvements in functioning" is erroneous. A WMD of 1.408 and $p<0.05$ indicate a statistically significant treatment effect of memantine versus placebo. The results of this restricted meta-analysis are close to the estimates in the meta-analyses of the six large RCTs.

PenTAG statement, p 217. "*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*"

Appendix O, page 315-319 of the Lundbeck submission (now published as Rive et al., 2010[1]) explains the methods and results of building the predictive equation of time to FTC. Both NPI total score and NPI domain score of hallucination were significant predictors of time to FTC in the univariate analysis (see Appendix O and the relevant extract below).

The NPI total score was selected by the stepwise procedure in the final model which considered all potential predictors, i.e. baseline and time-varying predictors. This model should be in all analyses as it provides more accurate predictions compared with an intermediate 'static' model. (The model shows lower AIC and higher R^2 as compared to the static model, see Appendix O).

The "static" model (including only baseline predictors) was built for the purpose of validation with external source, i.e. comparison with the predictive equation from the Assessment of health economics in Alzheimer's disease (AHEAD) model. The model cannot be regarded as final due to exclusion of time-varying predictors.

Indeed, the analysis shows that slopes of decline are strong predictors of time to FTC, and entered the model very early in the selection process, thereby modifying the entry criterion for other variables.

PenTAG statement, p 217. "*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."

As discussed in the major comments on the review of clinical effectiveness it is clearly inappropriate to use only the two RCTs in the moderately-severe to severe AD patient population, when all of the necessary data required to make an assessment across the full licensed indication for memantine was available in the Lundbeck submission.

It should also be noted that the statement underlined in the extract from page 217 for the TAR (underlined above) is incorrect as in fact the TAG's own meta-analysis included moderately severe to severe patients only.

PenTAG statement, p 217. "*The Reisberg and van Dyck study measured functional status using the ADCS-ADL19 (scores ranging between 0-54), not the ADCS-ADL23 (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."*

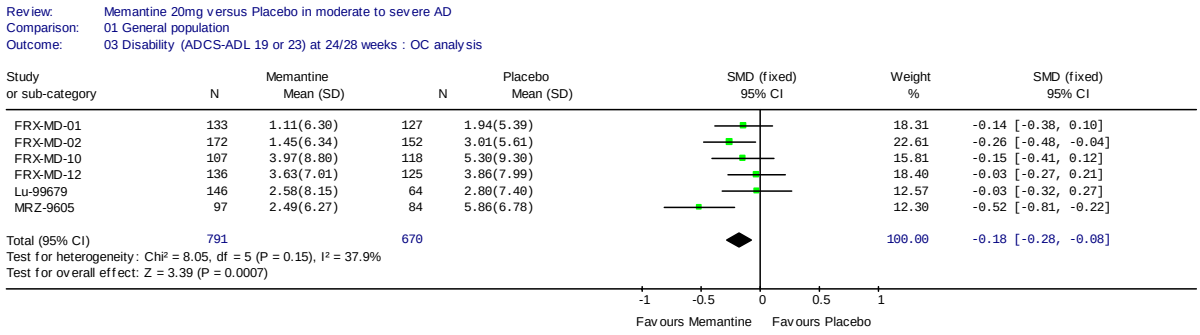
PenTAG statement, p 77. "*In this, account needs to be taken that the results in the Lundbeck submission are presented as SMD whereas those in the PenTAG analysis were WMDs. Approximate interconversion is achieved by multiplying or dividing by the pooled SD*."

In order to investigate the possible errors that might have been introduced by the rescaling of ADCS-ADL_{19} into ADCS-ADL_{23}, Lundbeck followed the PenTAG recommendation for interconversion between SMD and WMD. Rescaling ADCS-ADL_{19} into ADCS-ADL_{23} employing this approach produced similar estimates and thus validated the approach adopted by Lundbeck.

The approach consisted of the following steps:

- Using the meta-analysis on SMD pooling results from the ADCS-ADL_{19} and ADCS-ADL_{23} scales
- Estimating the pooled SD for each of these scales
- Multiplying these pooled SDs by the effect size to approximate the expected treatment effect on each scale
- Comparing the outcome of this analysis with meta-analysis on ADCS-ADL_{23} obtained by rescaling the ADCS-ADL_{19} for pooling

The figure below presents the results of the meta-analysis comparing the effect of memantine versus placebo on functional disability (pooled ADCS-ADL_{19} and ADCS-ADL_{23}) using SMD (this figure is extracted from dossier submitted by Lundbeck: Figure 2 from page 16 of the appendix).



For each study, SD was computed using the formula below (this is the same as that used by PenTAG, see equation 3 on page 68 of the TAR):

$$SD = \sqrt{\frac{(n_1 - 1) * S_1^2 + (n_2 - 1) * S_2^2}{n_1 + n_2 - 2}}$$

Where:

n_1, n_2: sample size in each treatment group

S_1, S_2: standard deviations associated with the mean change from baseline in each treatment group.

The table below shows the resulting pooled standard deviation across both groups, by study:

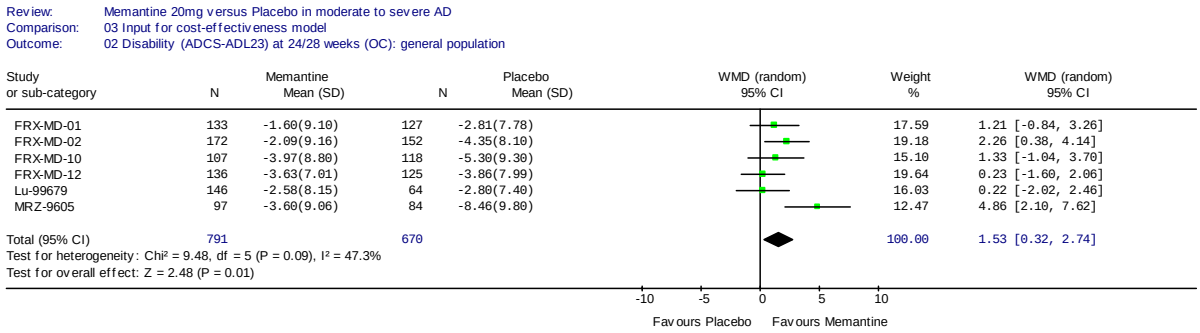
		Memantine			Placebo			Pooled
Study	**Scale**	**N**	**Mean**	**SD**	**N**	**Mean**	**SD**	**SD**
FRX-MD-01	ADCS-ADL$_{19}$	133	1.11	6.30	127	1.94	5.39	5.87
FRX-MD-02	ADCS-ADL$_{19}$	172	1.45	6.34	152	3.01	5.61	6.01
FRX-MD-10	ADCS-ADL$_{23}$	107	3.97	8.80	118	5.30	9.30	9.07
FRX-MD-12	ADCS-ADL$_{23}$	136	3.63	7.01	125	3.86	7.99	7.50
Lu-99679	ADCS-ADL$_{23}$	146	2.58	8.15	64	2.80	7.40	7.93
MRZ-9605	ADCS-ADL$_{19}$	97	2.49	6.27	84	5.86	6.78	6.51

The pooled SD for each scale was then simply estimated as the average of the pooled SD observed in all studies using that scale, weighted by the sample size of the study. This results in a pooled SD of 6.08 for ADCS-ADL_{19} and 8.13 for ADCS-ADL_{23}. When multiplying these by the SMD (0.18) from the meta-analysis:

- The benefit of memantine on ADCS-ADL_{19} is expected to be 1.09
- The benefit of memantine on ADCS-ADL_{23} is expected to be 1.46

The treatment benefit of memantine is expected to be higher on the ADCS-ADL_{23} compared to ADCS-ADL_{19}. This is due to the larger SDs for ADCS-ADL_{23}, which are expected to be related to the scale itself and not the study population as only sub-populations (moderate patients, i.e. MMSE 10-19) were considered in studies using ADCS-ADL_{23} while the full population (moderately severe to severe i.e. MMSE < 15) were considered in studies using ADCS-ADL_{19}.

When rescaling the ADCS-ADL_{19} (multiplying it by the ratio between its maximum and the maximum on the ADCS-ADL_{23}, i.e. 78/54) to pool results of studies using the ADCS-ADL_{19} with results of studies using the ADCS-ADL_{23}, the estimated treatment benefit of memantine was 1.53 (see figure below, reproduced from appendix N of the Lundbeck submission, page 265). This is very close to the estimate obtained using the PenTAG approach (1.46).



In conclusion, this alternative approach validates the rescaling of ADCS-ADL_{19} into ADCS-ADL_{23} by producing similar estimates.

MINOR COMMENTS

PenTAG statement, p 216. "*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*"

Lundbeck adopted the approach of the TA group during the previous 2004 evaluation. The model has been rebuilt to be appropriate for the memantine population, preserving all methodological assumptions. We propose to soften the wording and avoid implicit judgments.

The following wording is suggested "Specifically, changes on the ADAS-cog baseline, ADCS-ADL baseline and NPI baseline scores were meta-analysed and added to the related baseline variables in the risk equation."

The Probability of Death

MAJOR COMMENTS

PenTAG statement, p 217: "*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*".

In an attempt to answer the PenTAG concern regarding the death probability, an additional sensitivity analysis was conducted. To extreme, the probability of death regardless the model state was assumed to be twice as high as that in the base-case analysis.

The analysis showed that this chance had very small bearing on the model results and did not alter the model conclusion.

Results of one-way sensitivity analysis (doubled death rates)

	Cost (£, 2009)	QALYs	Time in pre-FTC
Memantine	£71,528	1.315	86.7 weeks
Standard care	£72,888	1.286	81.7weeks
Incremental	-£1,360	0.029	4.9 weeks

Under the assumption of this sensitivity analyses, proportion of patients reached 'death' state at the end of the 5-year evaluation period increased to 74.5% for both treatment strategies compared to 49.8% in the base case analysis. As expected, time in pre-FTC, QALYs and costs in both treatment arms were lower compared to those in base-case analysis. Nevertheless, compared to standard care memantine remained to be associated with a prolonged time in pre-FTC (4.9 weeks), QALYs gains (0.029) and cost-savings (£1,360). In corresponding stochastic analyses, the probability of memantine being more effective was 99.7% (99.8% in the base case analysis) and probability of memantine being less costly was 94.7% (96.4% in the base case analysis).

These results therefore fail to fully corroborate the PenTAG conclusion regarding the influence of mortality on results, indicating instead that even if outcomes per treatment strategy are indeed sensitive to the mortality rate, conclusions of the model in terms of difference between treatments remains stable, This is the case even when the mortality rate is drastically increased to double the base case.

PenTAG statement, p 218: "*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.*"

Clarification is needed on this statement "Receiving what?".

PenTAG statement, p 218: "*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*."

Lundbeck performed additional sensitivity analysis to investigate robustness of the model to this input variable. To extreme, the probability of death in FTC state was assumed to be twice as high as that in pre-FTC state, i.e. allowing patients in FTC state to leave the model sooner and incur less cost.

The analysis showed that this chance had very small bearing on the model results and did not alter the model conclusion.

Results of one-way sensitivity analysis (doubled death rates in FTC state)

	% dead	Cost (£, 2009)	QALYs	Time in pre-FTC
Memantine	70.5%	£79,404	1.416	91.4 weeks
Standard care	71.0%	£80,214	1.377	85.7 weeks
Incremental	-0.5%	-£810	0.038	5.6 weeks

Under the assumption of this sensitivity analyses, the proportion of patients reached 'death' state increased to around 71.0% at the end of the 5-year evaluation period as compared to around 49.8% in base case analysis. As expected, costs and QALYs for both treatment arms were lower as compared to those in base-case analysis. Nevertheless, compared to standard care memantine remained to be associated with a prolonged time in pre-FTC (5.6 weeks), QALYs gains (0.038) and cost-saving (-£810). In corresponding stochastic analysis, probability of memantine being more effective was 99.90% (99.75% in base case analysis) and probability of memantine being less costly was 85.43% (96.38% in base case analysis).

Costs

MAJOR COMMENTS

PenTAG statement, p 218. "*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 unit cost of memantine was taken from the BNF 58 (September 2009), the latest available issue at the time of submission (the BNF59 was issued in March 2010). The differences can be explained by the fact that Lundbeck considered the price at the ex-manufacturer level (i.e. sold to the wholesaler) while PenTAG considered the price sold to pharmacists. Furthermore TAG indicates that "this increased cost had little bearing on the base-case cost-effectiveness results". However the statement that "clearly its importance will be magnified if other changes are simultaneously made to the model, such as lessening the effect of memantine" is ambiguous and not supported by evidence.

Furthermore the TAR states, p 299: "*3 Monthly drug costs were calculated from costs reported in the BNF 58 for the specific doses of interest"* and *"Note that the relevant drug costs do not differ between BNF58 (4th quarter 2009) and BNF59 (1st quarter 2010)."*

PenTAG statement, p 218. "*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 statement should be amended to indicate that this was a conservative assumption.

The following wording is suggested:

"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. This may be deemed as a conservative assumption as for a targeted population initiation of treatment with memantine would not necessarily necessitate additional resource use".

PenTAG statement, p 218. "*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 LASER-AD study has been published in multiple peer-reviewed publications and the most relevant publications were submitted to NICE. For more information please see: Livingston et al., 2004; Ryu et al., 2005[14 28]

PenTAG statement, p 219." *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."*

The current estimates by Lundbeck of the cost per state are very close to the current cost estimate in the PenTAG model (please see the table below). The difference between the two is 11%. This is after the PenTAG model excluded all possible private payments (page 300 - Cost of health and social care received by Alzheimer's disease patients). In the sensitivity analyses conducted by Lundbeck on cost (variation of +/- 30%, which fully accounts for the difference in costs between the Lundbeck and PenTAG models), this did not change the model conclusions. This critique is unjustified and the statement should be removed.

Because PenTAG used an equation to estimate the cost per cycle in the 'pre-institution' state, we can compare only the cost of FTC in the Lundbeck model and cost of institution state in the PenTAG model.

Cost per 1 month, FTC state, Lundbeck	£3,267
Cost per 1 month, institution state, PenTAG, p309	£2,941

Please note that as stated in TAR on page 263:"*the previous model (SHTAC 2004 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*"

This feature was not implemented in the Lundbeck model. This is in fact a conservative approach for memantine as the cost decrease resulting from memantine health benefit could not be accounted for in patients who did not survive long enough to reach FTC. Clarification will be needed to make necessary adjustments and undertake further analyses.

PenTAG statement, p 219. "*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*."

Page numbers referred to the correct sections in PSSRU (not to Memantine submission), where all required information is clearly described (Curtis L. Unit Costs of Health and Social Care 2009, Personal Social Services Research Unit. 2009; Canterbury. University of Kent; 2009).

Utilities

MAJOR COMMENTS

PenTAG statement, p 220. "*Health benefits to individuals with AD were measured and valued within the analysis, but potential benefits to carers were not included*."

Not including health benefits to carers is in line with the standard methodology for economic evaluations. It should be noted that NICE rarely takes into consideration the cost and benefits of treatments to third parties). We will be willing to conduct additional analyses based on the TAR report to include those in the analyses.

PenTAG statement, p 220 "*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 Lundbeck submission presents sensitivity analyses employing the utility values from published sources Please see Appendix N, page 199.

The statement "*as direct EQ-5D scores were said to be absent" should be corrected. The EQ-5D values were not collected in the original study.*

PenTAG statement, p 220 "*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. 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*"

The mapping strategy is fully detailed in Appendix N. Additional requested information can be provided.

PenTAG statement, p 220. "*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)*"

This may be viewed as external validation of the mapping exercise. The submission presents sensitivity analyses employing the SHTAC utility values (see Appendix N, page 199). The model conclusion remained unchanged.

PenTAG statement, p 220 "*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."

Methods for model selection and the list of potential predictors are provided in Appendix N, page 170. Predictors other than ADCS-ADL were non significant after adjustment on ADCS-ADL.

PenTAG statement, p 220 "*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*".

This is an artificial problem. At 40 months, only 0.8% of memantine-treated patients (0.3% in standard care) are still in pre-FTC. An analysis was undertaken, in which a floor value (the estimated utility in FTC, i.e. 0.336) was used as a minimum for the utility in pre-FTC. This had no impact on results.

Results of one-way sensitivity analysis (a floor value of utility in FTC state as a minimum utility value in pre-FTC)

	Cost (£, 2009)	**QALYs**	**Time in pre-FTC**
Memantine	£93,076	1.533	91.4 weeks
Standard care	£94,787	1.502	85.7 weeks
Incremental	-£1,711	0.031	5.6 weeks

Adding a floor value for the utility in pre-FTC patients had no effect on the results of the model. Stochastic analysis also provided the same results as base case analysis i.e. a memantine is more effective with a probability of 99.75% and less costly with a probability of 96.38% compared to standard care.

PenTAG statement, p 220 **"***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.*"

This was done to address the limitation of the SHTAC 2004 model by allowing a gradual decrease of utility for patients in the pre-FTC state, in order to be able to also capture the benefit of the treatment in patients who did not survive long enough to reach FTC state. The same approach was used in the current TAR, page 263 "*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.*"

Extra sensitivity analysis on the General Population Base-Case

The role of this section is not clear. No accompanying explanation was provided. Please provide guidance on how the reader should treat this information.

Summary of memantine model comments

MAJOR COMMENTS: TBC

P221 "*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.

We believe that the main critiques of the Lundbeck model are not justified. Many of the requested 'missing' information were in fact presented in the Lundbeck dossier, appendices or references provided. The 50 page submission limit does not enable the manufacturer to present all the necessary details in a single document. We believe PenTAG could have taken more care in reviewing the submitted documentation. Furthermore, the major critiques of the TAR on the model submitted by Lundbeck are equally applicable to the PenTAG model itself (as argued above). A table has been complied to highlight this fact (Table 1 in Appendix B). We request that this statement is removed from the TAR.

P222 "*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.*"

As described in section 4 the remit of the TAR was to review the effectiveness of the AD therapies within their licensed indications. For memantine this is moderate to severe patients who are naive to AD medication, or have a history of past use of AChEIs or who are

currently on a stable dose of AChEIs. A meta-analysis of six large multicentre RCTs, which was conducted in accordance with internationally recognised standards (e.g. EMA and Cochrane) and has been widely published in the peer-reviewed literature, provides the evidence for the clinical efficacy of memantine in this licensed population. In fact, in 2005, the EMA extended the indication for memantine to moderate and severe AD on the basis of this meta-analysis. The approach to this meta-analysis in terms of pooling trials based on concomitant AChEI therapy was in line with the previous TAR for AD conducted in 2004. Despite the availability of this evidence the TAR did not consider the meta-analysis in their review and assessed only two of the six included trials.

The TAR review of memantine therefore does not cover the full licensed indication, although importantly this is not explicitly stated in the report.

Results – Memantine

MAJOR COMMENTS

The systematic searches conducted by the TAR went up to 31.03.2010 and therefore did not identify a cover a new economic evaluation of memantine in the UK published in May 2010.

The TAR commented *"The study by Tuomi175 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."*

It should be documented that this model for the UK case has now been presented in the public domain.

MINOR COMMENTS

On page 208 the TAR states that "*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*." The cited publications are for donepezil and galantamine and not memantine.

Appendix D: Clinical Trials for Emerging AD Therapies

Table 2: Overview of trials for emerging therapies in AD

Compound (INN)/ Expected indication	**Inclusion criteria for past/ concurrent pharmacological treatment**
Bapineuzumab Expected indication: mild-to-moderate AD	**Inclusion criteria for past/ concurrent pharmacological treatment** Concurrent use of cholinesterase inhibitor or memantine allowed, if stable • NCT00574132: Phase 3, Multicenter, Randomized, Double-Blind, Placebo-Controlled, Parallel-Group, Efficacy and Safety Trial of Bapineuzumab (AAB-001, ELN115727) In Patients With Mild to Moderate Alzheimer's Disease Who Are Apolipoprotein E4 Non- Carriers • NCT00575055: A Phase 3, Multicenter, Randomized, Double-Blind, Placebo-Controlled, Parallel-Group, Efficacy and Safety Trial of Bapineuzumab (AAB-001, ELN115727) In Patients With Mild to Moderate Alzheimer's Disease Who Are Apolipoprotein E4 Carriers. • NCT00676143: A Phase 3, Multicenter, Randomized, Double-Blind, Placebo-Controlled, Parallel-Group Efficacy and Safety Trial of Bapineuzumab in Subjects With Mild to Moderate Alzheimer Disease Who Are Apolipoprotein E e4 Carriers • NCT00667810: A Phase 3, Multicenter, Randomized, Double-Blind, Placebo-Controlled, Parallel-Group Efficacy and Safety Trial of Bapineuzumab in Subjects With Mild to Moderate Alzheimer Disease Who Are Apolipoprotein E e4 Non-Carriers

Gammagard Expected indication: mild-to-moderate AD	**Inclusion criteria for past / concurrent pharmacological treatment** Stable doses of approved AD medication(s) for at least 3 months prior to screening (e.g. AChE inhibitors, memantine): • NCT00299988 : A Placebo-Controlled, Randomized, Double-Blind Phase II Clinical Study of Gammagard Intravenous Immunoglobulin (IVIg) for Treatment of Mild to Moderate Alzheimer's Disease • NCT00812565: Prospective 24-Week, Double-Blind, Randomized, Multicenter, Placebo-Controlled Study Evaluating Safety and Change in Surrogate Parameters After Treatment With Increasing Dosages of Intravenous Immunoglobulin (IGIV) in Mild to Moderate Alzheimer's Disease

Dimebon Expected indication: mild-to-moderate AD in adjunct therapy to donepezil	**Trials with specified inclusion/exclusion criteria for ongoing/ concurrent pharmacological treatment** • NCT00838110: A Phase 3, Multi-Center, Randomized, Double-Blind Placebo-Controlled Study To Evaluate The Safety And Tolerability Of Dimebon (PF-01913539) For Up To 26-Weeks In Patients With Mild To Moderate Alzheimer's Disease **Inclusion criteria for ongoing/ concurrent pharmacological treatment** If on existing anti-dementia therapy, have been on a stable dose of anti-dementia therapy (cholinesterase inhibitors and/or memantine) for at least 60 days prior to dosing in study. If not taking existing anti-dementia therapy, have not received therapy with cholinesterase inhibitors and/or memantine within 60 days prior to dosing in this study. • NCT00829816: A Multi-Center Phase 1 Study of the Safety and Tolerability of Dimebon in Alzheimer's Disease Patients on Memantine (Cohort 1) and Memantine Plus Donepezil (Cohort 2) **Inclusion criteria for ongoing/ concurrent pharmacological treatment** Stable on memantine • NCT00912288; A Phase 3, Multi-Center, Randomized, Double-Blind, Placebo-Controlled 26-Week Trial To Evaluate The Efficacy And Safety Of **Dimebon** In Patients With Moderate-To-Severe Alzheimer's Disease **Inclusion criteria for ongoing/ concurrent pharmacological treatment:** Have been taking the medication memantine (ie., Namenda) for at least six months prior to this study. Exclusion Criteria: - Have taken medicines for Alzheimers disease other than memantine (e.g., donepezil, rivastigmine, galantamine, tacrine) within 2 months prior to this study. • NCT00954590: CONTACT: A Phase 3 Multicenter, Randomized, Double-Blind, Placebo-Controlled, Six-Month Safety and Efficacy Study of **Dimebon** in Patients With Moderate-to-Severe **Alzheimer's Disease** **Inclusion criteria for ongoing/ concurrent pharmacological treatment** Stable on donepezil for at least 6 months **Other trials (e.g. below) did not have specified exclusion/inclusion criteria for ongoing/ concurrent pharmacological treatment for AD** • NCT00675623: A Global Phase 3, Double-Blind, Placebo-Controlled Safety and Efficacy Study of Oral **Dimebon** in Patients With Mild-to-Moderate **Alzheimer's Disease** (CONNECTION) • NCT00829374: CONCERT: A Phase 3 Multicenter, Randomized, Placebo-Controlled, Double-Blind Twelve-Month Safety and Efficacy Study Evaluating **Dimebon** in Patients With Mild-to-Moderate **Alzheimer's Disease** on Donepezil • NCT00377715: Phase 2, Double-Blind, Placebo-Controlled Study of Oral **Dimebon** in Subjects With Mild to Moderate **Alzheimer's Disease**

Appendix E: CONSORT 2010 Checklist of Information Reported in Memantine Randomised Clinical Trials

CONSORT 2010 checklist of information to include when reporting a randomised trial*

Reisberg et al 2003 (MRZ-9605)

Section/Topic	Item No	Checklist item	Reported on page No
Title and abstract			
	1a	Identification as a randomised trial in the title	No, but stated in the Method section in the Abstract 1333
	1b	Structured summary of trial design, methods, results, and conclusions (for specific guidance see CONSORT for abstracts)	1333
Introduction			
Background and objectives	2a	Scientific background and explanation of rationale	1334
	2b	Specific objectives or hypotheses	1134
Methods			
Trial design	3a	Description of trial design (such as parallel, factorial) including allocation ratio	1334 METHODS/ Study Design Allocation ratio not specified (but numbers provided in e.g. Table 1)

			Study report gives full details
	3b	Important changes to methods after trial commencement (such as eligibility criteria), with reasons	Not mentioned. However, on p. 1334/Study Design it is stated "…and its ([trial]) amendments" Study report section 9.8 gives full details.
Participants	4a	Eligibility criteria for participants	1334
	4b	Settings and locations where the data were collected	No, but numbers of US centres listed in METHODS/ Study Design p. 1334, Col 2
Interventions	5	The interventions for each group with sufficient details to allow replication, including how and when they were actually administered	METHODS/ Study Design p. 1334, Col 2 (MEM 20mg pr. day or PBO). How and when not stated specifically
Outcomes	6a	Completely defined pre-specified primary and secondary outcome measures, including how and when they were assessed	1334 + 1335
	6b	Any changes to trial outcomes after the trial commenced, with reasons	Not stated
Sample size	7a	How sample size was determined	No
	7b	When applicable, explanation of any interim analyses and stopping guidelines	N/A
Randomisation:			

Sequence generation	8a	Method used to generate the random allocation sequence	1334 Study Design
	8b	Type of randomisation; details of any restriction (such as blocking and block size)	1334 Study Design Type of R not stated but block size included Study report section 9.4. gives full details.
Allocation concealment mechanism	9	Mechanism used to implement the random allocation sequence (such as sequentially numbered containers), describing any steps taken to conceal the sequence until interventions were assigned	No or 1334 Study Design Study report section 9.4 gives full details.
Implementation	10	Who generated the random allocation sequence, who enrolled participants, and who assigned participants to interventions	Not stated Study report section 9.4 gives full details.
Blinding	11a	If done, who was blinded after assignment to interventions (for example, participants, care providers, those assessing outcomes) and how	1334 Study Design "Staff at the individual sites blinded to the randomization process" Study report section 9.4 gives full details.
	11b	If relevant, description of the similarity of interventions	1334 Study Design
Statistical methods	12a	Statistical methods used to compare groups for primary and secondary outcomes	1336
	12b	Methods for additional analyses, such as subgroup analyses and	N/A

		adjusted analyses	
Results			
Participant flow (a diagram is strongly recommended)	13a	For each group, the numbers of participants who were randomly assigned, received intended treatment, and were analysed for the primary outcome	1336 and Table 2 1337
	13b	For each group, losses and exclusions after randomisation, together with reasons	1336 Col 2
Recruitment	14a	Dates defining the periods of recruitment and follow-up	No
	14b	Why the trial ended or was stopped	N/A – study completed as planned
Baseline data	15	A table showing baseline demographic and clinical characteristics for each group	1336 Table 1
Numbers analysed	16	For each group, number of participants (denominator) included in each analysis and whether the analysis was by original assigned groups	1337 Table 2
Outcomes and estimation	17a	For each primary and secondary outcome, results for each group, and the estimated effect size and its precision (such as 95% confidence interval)	1337 Table 2
	17b	For binary outcomes, presentation of both absolute and relative effect sizes is recommended	N/A
Ancillary analyses	18	Results of any other analyses performed, including subgroup analyses and adjusted analyses, distinguishing pre-specified from exploratory	1337-1339 distinguishing pre-specified from exploratory is not indicated
Harms	19	All important harms or unintended effects in each group (for specific	1339

		guidance see CONSORT for harms)	
Discussion			
Limitations	20	Trial limitations, addressing sources of potential bias, imprecision, and, if relevant, multiplicity of analyses	1339-1340
Generalisability	21	Generalisability (external validity, applicability) of the trial findings	1339-1340
Interpretation	22	Interpretation consistent with results, balancing benefits and harms, and considering other relevant evidence	1339-1340
Other information			
Registration	23	Registration number and name of trial registry	N/A
Protocol	24	Where the full trial protocol can be accessed, if available	N/A
Funding	25	Sources of funding and other support (such as supply of drugs), role of funders	1334 and 1340

*We strongly recommend reading this statement in conjunction with the CONSORT 2010 Explanation and Elaboration for important clarifications on all the items. If relevant, we also recommend reading CONSORT extensions for cluster randomised trials, non-inferiority and equivalence trials, non-pharmacological treatments, herbal interventions, and pragmatic trials. Additional extensions are forthcoming: for those and for up to date references relevant to this checklist, see www.consort-statement.org.

CONSORT 2010 checklist of information to include when reporting a randomised trial*

Tariot et al 2004 (MD02)

Section/Topic	Item No	Checklist item	Reported on page No
Title and abstract			
	1a	Identification as a randomised trial in the title	317
	1b	Structured summary of trial design, methods, results, and conclusions (for specific guidance see CONSORT for abstracts)	317
Introduction			
Background and objectives	2a	Scientific background and explanation of rationale	317
	2b	Specific objectives or hypotheses	317
Methods			
Trial design	3a	Description of trial design (such as parallel, factorial) including allocation ratio	318 Allocation ratio not specified (but numbers provided in e.g. Fig. 1) Study report section 5.3 gives full details
	3b	Important changes to methods after trial commencement (such as eligibility criteria), with reasons	N/A
Participants	4a	Eligibility criteria for participants	318 METHODS/

			Participants
	4b	Settings and locations where the data were collected	No, but number of US sites is mentioned on p. 318 METHODS/ Participants
Interventions	5	The interventions for each group with sufficient details to allow replication, including how and when they were actually administered	318 Interventions How and when not stated specifically Study report section 5.3 gives full details
Outcomes	6a	Completely defined pre-specified primary and secondary outcome measures, including how and when they were assessed	318-319 Outcome Measures
	6b	Any changes to trial outcomes after the trial commenced, with reasons	Not stated
Sample size	7a	How sample size was determined	319 Sample Size
	7b	When applicable, explanation of any interim analyses and stopping guidelines	N/A
Randomisation:			
Sequence generation	8a	Method used to generate the random allocation sequence	318 Interventions ("…randomization list generated and retained") Study report section 5.3 gives full details

	8b	Type of randomisation; details of any restriction (such as blocking and block size)	318 Interventions Type of R not stated but block size included Study report section 5.3 gives full details
Allocation concealment mechanism	9	Mechanism used to implement the random allocation sequence (such as sequentially numbered containers), describing any steps taken to conceal the sequence until interventions were assigned	318 Interventions "At baseline visit, each INV sequentially assigned a randomization number to each patient. No individual pt R code was revealed during the trial" Study report section 5.3 gives full details
Implementation	10	Who generated the random allocation sequence, who enrolled participants, and who assigned participants to interventions	Not stated Study report section 5.3 gives full details
Blinding	11a	If done, who was blinded after assignment to interventions (for example, participants, care providers, those assessing outcomes) and how	Double-.blind study design; thus all were blinded
	11b	If relevant, description of the similarity of interventions	318 METHODS/ Interventions
Statistical methods	12a	Statistical methods used to compare groups for primary and secondary outcomes	319 METHODS/ Stat Analyses
	12b	Methods for additional analyses, such as subgroup analyses and	319 METHODS/

		adjusted analyses	Stat Analyses
Results			
Participant flow (a diagram is strongly recommended)	13a	For each group, the numbers of participants who were randomly assigned, received intended treatment, and were analysed for the primary outcome	320 Figure 1
	13b	For each group, losses and exclusions after randomisation, together with reasons	320 Figure 1
Recruitment	14a	Dates defining the periods of recruitment and follow-up	318 METHODS/ Participants
	14b	Why the trial ended or was stopped	N/A – study completed as planned
Baseline data	15	A table showing baseline demographic and clinical characteristics for each group	321 Table 1
Numbers analysed	16	For each group, number of participants (denominator) included in each analysis and whether the analysis was by original assigned groups	319 METHODS/ Stat Anal
Outcomes and estimation	17a	For each primary and secondary outcome, results for each group, and the estimated effect size and its precision (such as 95% confidence interval)	321 Table 2
	17b	For binary outcomes, presentation of both absolute and relative effect sizes is recommended	N/A
Ancillary analyses	18	Results of any other analyses performed, including subgroup analyses and adjusted analyses, distinguishing pre-specified from exploratory	320 Prim outcome 319 Stat anal

Harms	19	All important harms or unintended effects in each group (for specific guidance see CONSORT for harms)	Potential SAEs not included
Discussion			
Limitations	20	Trial limitations, addressing sources of potential bias, imprecision, and, if relevant, multiplicity of analyses	323 Col 2
Generalisability	21	Generalisability (external validity, applicability) of the trial findings	323 Col 2
Interpretation	22	Interpretation consistent with results, balancing benefits and harms, and considering other relevant evidence	323
Other information			
Registration	23	Registration number and name of trial registry	MD02 on forestclinicaltrials.com/CTR
Protocol	24	Where the full trial protocol can be accessed, if available	Synopsis available on forestclinicaltrials.com/CTR
Funding	25	Sources of funding and other support (such as supply of drugs), role of funders	324

*We strongly recommend reading this statement in conjunction with the CONSORT 2010 Explanation and Elaboration for important clarifications on all the items. If relevant, we also recommend reading CONSORT extensions for cluster randomised trials, non-inferiority and equivalence trials, non-pharmacological treatments, herbal interventions, and pragmatic trials. Additional extensions are forthcoming: for those and for up to date references relevant to this checklist, see www.consort-statement.org.

CONSORT 2010 checklist of information to include when reporting a randomised trial*

Porsteinsson et al 2008 (MD12)

Section/Topic	Item No	Checklist item	Reported on page No
Title and abstract			
	1a	Identification as a randomised trial in the title	83
	1b	Structured summary of trial design, methods, results, and conclusions (for specific guidance see CONSORT for abstracts)	83
Introduction			
Background and objectives	2a	Scientific background and explanation of rationale	83 Hypotheses rely on clinical data
	2b	Specific objectives or hypotheses	83 Hypotheses rely on clinical data
Methods			
Trial design	3a	Description of trial design (such as parallel, factorial) including allocation ratio	84 Allocation ratio not specified (but numbers provided in e.g. Fig. 1) Study report section 9.4. gives full details
	3b	Important changes to methods after trial commencement (such as eligibility	N/A

		criteria), with reasons	
Participants	4a	Eligibility criteria for participants	84 METHODS/ Participants
	4b	Settings and locations where the data were collected	83 METHODS/ Participants
Interventions	5	The interventions for each group with sufficient details to allow replication, including how and when they were actually administered	84 METHODS/ Interventions
Outcomes	6a	Completely defined pre-specified primary and secondary outcome measures, including how and when they were assessed	84 METHODS/ Outcome Measures
	6b	Any changes to trial outcomes after the trial commenced, with reasons	Not stated
Sample size	7a	How sample size was determined	84 METHODS/ Statistical Analyses
	7b	When applicable, explanation of any interim analyses and stopping guidelines	N/A 84 METHODS/ Statistical Analyses "No IA were performed"

Randomisation:			
Sequence generation	8a	Method used to generate the random allocation sequence	84
	8b	Type of randomisation; details of any restriction (such as blocking and block size)	84 Type of R not stated but block size included Study report section 9.4. gives full details
Allocation concealment mechanism	9	Mechanism used to implement the random allocation sequence (such as sequentially numbered containers), describing any steps taken to conceal the sequence until interventions were assigned	Study report section 9.4. gives full details
Implementation	10	Who generated the random allocation sequence, who enrolled participants, and who assigned participants to interventions	Not stated Study report section 9.4. gives full details
Blinding	11a	If done, who was blinded after assignment to interventions (for example, participants, care providers, those assessing outcomes) and how	Double-.blind study design; thus all were blinded Study report section 9.4. gives full details
	11b	If relevant, description of the similarity of interventions	84 METHODS/ Interventions
Statistical methods	12a	Statistical methods used to compare groups for primary and secondary outcomes	84 METHODS/ Statistical Analyses

	12b	Methods for additional analyses, such as subgroup analyses and adjusted analyses	84 METHODS/ Statistical Analyses
Results			
Participant flow (a diagram is strongly recommended)	13a	For each group, the numbers of participants who were randomly assigned, received intended treatment, and were analysed for the primary outcome	85 Fig 1
	13b	For each group, losses and exclusions after randomisation, together with reasons	85 Fig 1
Recruitment	14a	Dates defining the periods of recruitment and follow-up	83 METHODS/ Participants
	14b	Why the trial ended or was stopped	N/A – study completed as planned
Baseline data	15	A table showing baseline demographic and clinical characteristics for each group	86 Table 1
Numbers analysed	16	For each group, number of participants (denominator) included in each analysis and whether the analysis was by original assigned groups	87 Table 2
Outcomes and estimation	17a	For each primary and secondary outcome, results for each group, and the estimated effect size and its precision (such as 95% confidence interval)	87 Table 2 and 3
	17b	For binary outcomes, presentation of both absolute and relative effect sizes is recommended	N/A
Ancillary analyses	18	Results of any other analyses performed, including subgroup analyses and adjusted analyses, distinguishing pre-specified from exploratory	N/A

Harms	19	All important harms or unintended effects in each group (for specific guidance see CONSORT for harms)	85 and Table 4
Discussion			
Limitations	20	Trial limitations, addressing sources of potential bias, imprecision, and, if relevant, multiplicity of analyses	87-88
Generalisability	21	Generalisability (external validity, applicability) of the trial findings	87-88
Interpretation	22	Interpretation consistent with results, balancing benefits and harms, and considering other relevant evidence	87-88
Other information			
Registration	23	Registration number and name of trial registry	MD12 on forestclinicaltrials.com/CTR
Protocol	24	Where the full trial protocol can be accessed, if available	Synopsis available on www.forestclinicaltrials.com
Funding	25	Sources of funding and other support (such as supply of drugs), role of funders	88

*We strongly recommend reading this statement in conjunction with the CONSORT 2010 Explanation and Elaboration for important clarifications on all the items. If relevant, we also recommend reading CONSORT extensions for cluster randomised trials, non-inferiority and equivalence trials, non-pharmacological treatments, herbal interventions, and pragmatic trials. Additional extensions are forthcoming: for those and for up to date references relevant to this checklist, see www.consort-statement.org.

CONSORT 2010 checklist of information to include when reporting a randomised trial*

Van Dyck et al 2007 (MDO1)

Section/Topic	Item No	Checklist item	Reported on page No
Title and abstract			
	1a	Identification as a randomised trial in the title	136
	1b	Structured summary of trial design, methods, results, and conclusions (for specific guidance see CONSORT for abstracts)	136
Introduction			
Background and objectives	2a	Scientific background and explanation of rationale	136
	2b	Specific objectives or hypotheses	136
Methods			
Trial design	3a	Description of trial design (such as parallel, factorial) including allocation ratio	136 Participants Allocation ratio not specified (but numbers provided in e.g. Figure 1) Study report section 9.4. gives full details
	3b	Important changes to methods after trial commencement (such as eligibility criteria), with reasons	Not mentioned.

Participants	4a	Eligibility criteria for participants	136-137 Participants
	4b	Settings and locations where the data were collected	No, but numbers of US sites are provided
Interventions	5	The interventions for each group with sufficient details to allow replication, including how and when they were actually administered	137 Interventions
Outcomes	6a	Completely defined pre-specified primary and secondary outcome measures, including how and when they were assessed	137 Outcome Measures
	6b	Any changes to trial outcomes after the trial commenced, with reasons	Not stated
Sample size	7a	How sample size was determined	137 Stat Analyses
	7b	When applicable, explanation of any interim analyses and stopping guidelines	137 Stat Analyses "No interim analyses were performed"
Randomisation:			
Sequence generation	8a	Method used to generate the random allocation sequence	Not stated Study report section 9.4 gives full details
	8b	Type of randomisation; details of any restriction (such as blocking and block size)	Not stated Study report section 9.4 gives full details
Allocation concealment	9	Mechanism used to implement the random allocation sequence (such as sequentially numbered containers), describing any steps taken to conceal the sequence until	Not stated Study report section 9.4 gives full details

mechanism		interventions were assigned	
Implementation	10	Who generated the random allocation sequence, who enrolled participants, and who assigned participants to interventions	Not stated Study report section 9.4 gives full details
Blinding	11a	If done, who was blinded after assignment to interventions (for example, participants, care providers, those assessing outcomes) and how	Double-.blind study design; thus all were blinded Study report section 9.4 gives full details
	11b	If relevant, description of the similarity of interventions	Not stated Study report section 9.4 gives full details
Statistical methods	12a	Statistical methods used to compare groups for primary and secondary outcomes	137+138
	12b	Methods for additional analyses, such as subgroup analyses and adjusted analyses	138
Results			
Participant flow (a diagram is strongly recommended)	13a	For each group, the numbers of participants who were randomly assigned, received intended treatment, and were analysed for the primary outcome	138 Fig 1
	13b	For each group, losses and exclusions after randomisation, together with reasons	138 Fig 1
Recruitment	14a	Dates defining the periods of recruitment and follow-up	136 Participants
	14b	Why the trial ended or was stopped	N/A – study completed as planned
Baseline data	15	A table showing baseline demographic and clinical characteristics for each group	139

			Table 1
Numbers analysed	16	For each group, number of participants (denominator) included in each analysis and whether the analysis was by original assigned groups	139 Table 2
Outcomes and estimation	17a	For each primary and secondary outcome, results for each group, and the estimated effect size and its precision (such as 95% confidence interval)	139 Table 2
	17b	For binary outcomes, presentation of both absolute and relative effect sizes is recommended	N/A
Ancillary analyses	18	Results of any other analyses performed, including subgroup analyses and adjusted analyses, distinguishing pre-specified from exploratory	140 Post-hoc Analyses
Harms	19	All important harms or unintended effects in each group (for specific guidance see CONSORT for harms)	140-141 Safety and Tolerability
Discussion			
Limitations	20	Trial limitations, addressing sources of potential bias, imprecision, and, if relevant, multiplicity of analyses	141-142
Generalisability	21	Generalisability (external validity, applicability) of the trial findings	141-142
Interpretation	22	Interpretation consistent with results, balancing benefits and harms, and considering other relevant evidence	141-142
Other information			
Registration	23	Registration number and name of trial registry	MD01 on forestclinicaltrials.com/CTR
Protocol	24	Where the full trial protocol can be accessed, if available	Synopsis available on forestclinicaltrials.com/CTR

Funding	25	Sources of funding and other support (such as supply of drugs), role of funders	136

*We strongly recommend reading this statement in conjunction with the CONSORT 2010 Explanation and Elaboration for important clarifications on all the items. If relevant, we also recommend reading CONSORT extensions for cluster randomised trials, non-inferiority and equivalence trials, non-pharmacological treatments, herbal interventions, and pragmatic trials. Additional extensions are forthcoming: for those and for up to date references relevant to this checklist, see www.consort-statement.org.

CONSORT 2010 checklist of information to include when reporting a randomised trial*

Peskind et al 2006 (MD10)

Section/Topic	Item No	Checklist item	Reported on page No
Title and abstract			
	1a	Identification as a randomised trial in the title	704
	1b	Structured summary of trial design, methods, results, and conclusions (for specific guidance see CONSORT for abstracts)	704
Introduction			
Background and objectives	2a	Scientific background and explanation of rationale	705
	2b	Specific objectives or hypotheses	705
Methods			
Trial design	3a	Description of trial design (such as parallel, factorial) including allocation ratio	706 METHODS/ Interventions Allocation ratio not specified (but numbers provided in e.g. Fig. 1) Study report section 9.4. gives full details
	3b	Important changes to methods after trial commencement (such as eligibility criteria), with reasons	N/A

Participants	4a	Eligibility criteria for participants	705 METHODS/ Participants
	4b	Settings and locations where the data were collected	705 METHODS/ Participants (numbers of US sites are provided)
Interventions	5	The interventions for each group with sufficient details to allow replication, including how and when they were actually administered	706 METHODS/ Interventions
Outcomes	6a	Completely defined pre-specified primary and secondary outcome measures, including how and when they were assessed	706 METHODS/ Outcome Measures
	6b	Any changes to trial outcomes after the trial commenced, with reasons	Not stated
Sample size	7a	How sample size was determined	706 METHODS/ Sample Size
	7b	When applicable, explanation of any interim analyses and stopping guidelines	N/A 707 METHODS/ Statistical Analyses "No IA were performed"
Randomisation:			

Sequence generation	8a	Method used to generate the random allocation sequence	Not provided Study report section 9.4 gives full details
	8b	Type of randomisation; details of any restriction (such as blocking and block size)	706 Type of R not stated but block size included Study report section 9.4 gives full details
Allocation concealment mechanism	9	Mechanism used to implement the random allocation sequence (such as sequentially numbered containers), describing any steps taken to conceal the sequence until interventions were assigned	Not provided Study report section 9.4 gives full details
Implementation	10	Who generated the random allocation sequence, who enrolled participants, and who assigned participants to interventions	Not provided Study report section 9.4 gives full details
Blinding	11a	If done, who was blinded after assignment to interventions (for example, participants, care providers, those assessing outcomes) and how	Double-.blind study design; thus all were blinded Study report section 9.4 gives full details
	11b	If relevant, description of the similarity of interventions	706 METHODS/ Interventions
Statistical methods	12a	Statistical methods used to compare groups for primary and secondary outcomes	707 METHODS/ Statistical Analyses

	12b	Methods for additional analyses, such as subgroup analyses and adjusted analyses	707 METHODS/ Statistical Analyses
Results			
Participant flow (a diagram is strongly recommended)	13a	For each group, the numbers of participants who were randomly assigned, received intended treatment, and were analysed for the primary outcome	707 Fig 1
	13b	For each group, losses and exclusions after randomisation, together with reasons	707 Fig 1
Recruitment	14a	Dates defining the periods of recruitment and follow-up	705 METHODS/ Participants
	14b	Why the trial ended or was stopped	N/A – study completed as planned
Baseline data	15	A table showing baseline demographic and clinical characteristics for each group	708 Table 1
Numbers analysed	16	For each group, number of participants (denominator) included in each analysis and whether the analysis was by original assigned groups	709 Table 2
Outcomes and estimation	17a	For each primary and secondary outcome, results for each group, and the estimated effect size and its precision (such as 95% confidence interval)	709 Table 2 710-711 Figs 2-4
	17b	For binary outcomes, presentation of both absolute and relative effect sizes is recommended	N/A
Ancillary analyses	18	Results of any other analyses performed, including subgroup analyses and adjusted analyses, distinguishing pre-specified from exploratory	N/A

Harms	19	All important harms or unintended effects in each group (for specific guidance see CONSORT for harms)	709-711 and Table 3
Discussion			
Limitations	20	Trial limitations, addressing sources of potential bias, imprecision, and, if relevant, multiplicity of analyses	711-713
Generalisability	21	Generalisability (external validity, applicability) of the trial findings	711-713
Interpretation	22	Interpretation consistent with results, balancing benefits and harms, and considering other relevant evidence	711-713
Other information			
Registration	23	Registration number and name of trial registry	N/A
Protocol	24	Where the full trial protocol can be accessed, if available	N/A
Funding	25	Sources of funding and other support (such as supply of drugs), role of funders	713-714

*We strongly recommend reading this statement in conjunction with the CONSORT 2010 Explanation and Elaboration for important clarifications on all the items. If relevant, we also recommend reading CONSORT extensions for cluster randomised trials, non-inferiority and equivalence trials, non-pharmacological treatments, herbal interventions, and pragmatic trials. Additional extensions are forthcoming: for those and for up to date references relevant to this checklist, see www.consort-statement.org.

CONSORT 2010 checklist of information to include when reporting a randomised trial*

Bakchine & Loft 2008 (Lu-99679)

Section/Topic	Item No	Checklist item	Reported on page No
Title and abstract			
	1a	Identification as a randomised trial in the title	97
	1b	Structured summary of trial design, methods, results, and conclusions (for specific guidance see CONSORT for abstracts)	97
Introduction			
Background and objectives	2a	Scientific background and explanation of rationale	97-98
	2b	Specific objectives or hypotheses	97-98
Methods			
Trial design	3a	Description of trial design (such as parallel, factorial) including allocation ratio	98 MATERIALS AND METHODS/ Study design
	3b	Important changes to methods after trial commencement (such as eligibility criteria), with reasons	N/A
Participants	4a	Eligibility criteria for participants	98 MATERIALS AND METHODS/ Patients

	4b	Settings and locations where the data were collected	705 METHODS/ Study design
Interventions	5	The interventions for each group with sufficient details to allow replication, including how and when they were actually administered	98 MATERIALS AND METHODS/ Study design
Outcomes	6a	Completely defined pre-specified primary and secondary outcome measures, including how and when they were assessed	99 MATERIALS AND METHODS / Primary and Secondary efficacy variables 100 Primary and secondary efficacy assessments
	6b	Any changes to trial outcomes after the trial commenced, with reasons	Not stated
Sample size	7a	How sample size was determined	99 MATERIALS AND METHODS / Power and sample size calculations
	7b	When applicable, explanation of any interim analyses and stopping guidelines	N/A 100 MATERIALS AND METHODS / Secondary eff assessments "No IA were performed"
Randomisation:			

Sequence generation	8a	Method used to generate the random allocation sequence	Not provided Study report section 5.3 gives full details
	8b	Type of randomisation; details of any restriction (such as blocking and block size)	Not provided Study report section 5.3 gives full details
Allocation concealment mechanism	9	Mechanism used to implement the random allocation sequence (such as sequentially numbered containers), describing any steps taken to conceal the sequence until interventions were assigned	Not provided Study report section 5.3 gives full details
Implementation	10	Who generated the random allocation sequence, who enrolled participants, and who assigned participants to interventions	Not provided Study report section 5.3 gives full details
Blinding	11a	If done, who was blinded after assignment to interventions (for example, participants, care providers, those assessing outcomes) and how	100 MATERIALS AND METHODS / Allocation to treatment "All study personnel and participants were blinded to treatment assignment for the duration of the study"
	11b	If relevant, description of the similarity of interventions	100 MATERIALS AND METHODS/ Allocation to treatment
Statistical methods	12a	Statistical methods used to compare groups for primary and secondary outcomes	100 MATERIALS AND METHODS/ Primary and secondary efficacy assessments

	12b	Methods for additional analyses, such as subgroup analyses and adjusted analyses	N/A
Results			
Participant flow (a diagram is strongly recommended)	13a	For each group, the numbers of participants who were randomly assigned, received intended treatment, and were analysed for the primary outcome	102 RESULTS Fig 1
	13b	For each group, losses and exclusions after randomisation, together with reasons	102 RESULTS Fig 1
Recruitment	14a	Dates defining the periods of recruitment and follow-up	98 MATERIALS AND METHODS/ Participants
	14b	Why the trial ended or was stopped	N/A – study completed as planned
Baseline data	15	A table showing baseline demographic and clinical characteristics for each group	101 Table 1
Numbers analysed	16	For each group, number of participants (denominator) included in each analysis and whether the analysis was by original assigned groups	102 Fig 1 Table 2
Outcomes and estimation	17a	For each primary and secondary outcome, results for each group, and the estimated effect size and its precision (such as 95% confidence interval)	101-104 Table 2+3 Fig 2-3
	17b	For binary outcomes, presentation of both absolute and relative effect sizes is recommended	N/A
Ancillary analyses	18	Results of any other analyses performed, including subgroup analyses and adjusted	N/A

		analyses, distinguishing pre-specified from exploratory	
Harms	19	All important harms or unintended effects in each group (for specific guidance see CONSORT for harms)	101-103 Table 4
Discussion			
Limitations	20	Trial limitations, addressing sources of potential bias, imprecision, and, if relevant, multiplicity of analyses	103-105
Generalisability	21	Generalisability (external validity, applicability) of the trial findings	103-105
Interpretation	22	Interpretation consistent with results, balancing benefits and harms, and considering other relevant evidence	103-105
Other information			
Registration	23	Registration number and name of trial registry	99679 on www.lundbecktrials.com
Protocol	24	Where the full trial protocol can be accessed, if available	Summary available on Lundbecktrials.com
Funding	25	Sources of funding and other support (such as supply of drugs), role of funders	105

*We strongly recommend reading this statement in conjunction with the CONSORT 2010 Explanation and Elaboration for important clarifications on all the items. If relevant, we also recommend reading CONSORT extensions for cluster randomised trials, non-inferiority and equivalence trials, non-pharmacological treatments, herbal interventions, and pragmatic trials. Additional extensions are forthcoming: for those and for up to date references relevant to this checklist, see www.consort-statement.org.

Appendix F: Response to PenTAG Critique of Memantine Randomised Clinical Trials

TAR Quality appraisal for Porsteinsson et al. (2008) – MD12 –Appendix 3			
	Items of the TAR Quality Appraisal	**Classification according to TAR**	**Lundbeck comments for items classified by PenTAG as inadequate or unknown**
1	**Was the assignment to the treatment groups really random?**	ADEQUATE	
2	**Was the treatment allocation concealed?**	ADEQUATE	
3	**Were the groups similar at baseline in terms of prognostic factors?**	REPORTED - YES	
4	**Were the eligibility criteria specified?**	INADEQUATE	Disagree. Eligibility criteria specified on page 84 METHODS/Participants
5	**Were outcome assessors blinded to the treatment allocation?**	UNKNOWN	Disagree. The manuscript explains that this was a double-blind study design; thus all were blinded. Furthermore, clinical study report (section 9.4) confirms that no double-blind treatment assignment was unblinded before database lock thus all remained blinded until the end of the study.
6	**Was the care provider blinded?**	ADEQUATE	
7	**Was the patient blinded?**	ADEQUATE	

8	**Were the point estimates and measure of variability presented for the primary outcome measure?**	ADEQUATE	
9	**Did the analyses include an intention-to-treat analysis?**	ADEQUATE	
10	**Were withdrawals and dropouts completely described?**	ADEQUATE	

TAR Quality appraisal for Van Dyck et al. (2007)– MD01 –Appendix 3			
	Items of the TAR Quality Appraisal	**Classification according to TAR**	**Lundbeck comments for items classified by PenTAG as inadequate or partial**
1	**Was the assignment to the treatment groups really random?**	UNKNOWN	The manuscript mentions that the study was approved by IRBs and so this is testimony that procedures relating to random assignment would have been adequate. Furthermore, section 9.4 of the clinical study report confirms that patients who met all of the eligibility criteria at baseline were randomized (1:1 ratio) to one of two treatment groups, either memantine or placebo, in accordance with the randomization list generated by the Sponsor's Statistical Programming Department. Each study site was provided with double-blind drug supplies corresponding to a single sequence of patient randomization numbers. At the Baseline visit, patients were sequentially assigned the randomization numbers.
2	**Was the treatment allocation concealed?**	UNKNOWN	The manuscript mentions that the study was approved by IRBs and so this is testimony that procedures relating to treatment allocation would have been adequate. Section 9.4 of the clinical study report confirms that a hard copy of the randomization list was retained by the Sponsor's Department of Drug Safety Surveillance, in a secure, locked area. Treatment codes were unblinded at the termination of the study after the database was locked.
3	**Were the groups similar at baseline in terms of prognostic factors?**	REPORTED - YES	

4	**Were the eligibility criteria specified?**	UNKNOWN	Disagree. Eligibility criteria is specified on pages136-137 METHODS/Participants Participants
5	**Were outcome assessors blinded to the treatment allocation?**	UNKNOWN	Disagree. The manuscript explains that this was a double-blind study design; thus all were blinded. The manuscript also mentions that the study was approved by IRBs and so this is testimony that procedures relating to blinding would have been adequate. Furthermore, clinical study report (section 9.4) confirms that no double-blind treatment assignment was unblinded before database thus all remained blinded until the end of the study.
6	**Was the care provider blinded?**	PARTIAL	Disagree. The manuscript explains that this was a double-blind study design; thus all were blinded. The manuscript also mentions that the study was approved by IRBs and so this is testimony that procedures relating to blinding would have been adequate. Furthermore, clinical study report (section 9.4) confirms that no double-blind treatment assignment was unblinded before database lock thus all remained blinded until the end of the study.
7	**Was the patient blinded?**	PARTIAL	Diasagree. The manuscript explains that this was a double-blind study design; thus all were blinded. The manuscript also mentions that the study was approved by IRBs and so this is testimony that procedures relating to blinding would have been adequate. Furthermore, clinical study report (section 9.4) confirms that no double-blind treatment assignment was unblinded before database lock thus all remained blinded until the end of the study.

8	**Were the point estimates and measure of variability presented for the primary outcome measure?**	ADEQUATE	
9	**Did the analyses include an intention-to-treat analysis?**	ADEQUATE	
10	**Were withdrawals and dropouts completely described?**	ADEQUATE	

References

[1] Rive B, Le Reun C, Grishchenko M, et al., 2010. Predicting time to full-time care in AD: a new model. JME, 13(2): 362-70.

[2] Caro JJ, Getsios D, Migliaccio-Walle K, et al. Assessment of health economics in Alzheimer's disease (AHEAD) based on need for fulltime care. Neurology 2001; 57: 964–971.

[3] Wolstenholme J, Fenn P, Gray A, Keene J, Jacoby R, Hope T. Estimating the relationship between disease progression and cost of care in dementia. Br J Psychiatry. 2002 Jul;181:36-42

[4] Kurz X, Scuvee-Moreau J, Rive B, et al. A new approach to the qualitative evaluation of functional disability in dementia. Int J Geriatr Psychiatry 2003 Nov;18(11):1050-5

[5] Winblad B, Jones RW, Wirth Y, Stöffler A, Möbius HJ. Memantine in moderate to severe Alzheimer's disease: a meta-analysis of randomised clinical trials. Dement Geriatr Cogn Disord. 2007; 24(1):20-7.

[6] Mecocci P, Bladström A, Stender K. Effects of memantine on cognition in patients with moderate to severe Alzheimer's disease: post-hoc analyses of ADAS-cog and SIB total and single-item scores from six randomized, double-blind, placebo-controlled studies. Int J Geriatr Psychiatry. 2009 May;24(5):532-8.

[7] Gauthier S, Loft H, Cummings J. Improvement in behavioural symptoms in patients with moderate to severe Alzheimer's disease by memantine: a pooled data analysis. Int J Geriatr Psychiatry. 2008 May;23(5):537-45.

[8] Winblad B et al. J Nutr Health Aging 2010

[9] Press Release from Elan Corporation, 13 January 2009

[10] Press Release from Medivation, Jan 28, 2008

[11]EMEA guidelines on medicinal products for the treatment of AD and other dementia http://www.tga.gov.au/docs/pdf/euguide/ewp/055395enrev1.pdf

[12] McShane R, Areosa Sastre A, Minakaran N. Memantine for dementia. Cochrane Database Syst Rev 2006(2):CD003154

[13] Vellas B, Andrieu S, Sampaio C, Coley N, Wilcock G; European Task Force Group. Endpoints for trials in Alzheimer's disease: a European task force consensus. Lancet Neurol. 2008 May;7(5):436-50. Review

[14] Ryu SH, Katona C, Rive B, et al. Persistence of and changes in neuropsychiatric symptoms in Alzheimer disease over 6 months: the LASER-AD study. Am J Geriatr Psychiatry 2005 Nov;13(11):976-83

[15] Wragg RE, Jeste DV. Overview of depression and psychosis in Alzheimer's disease. Am J Psychiatry 1989 May;146(5):577-87

[16] Ropacki SA, Jeste DV. Epidemiology of and risk factors for psychosis of Alzheimer's disease: a review of 55 studies published from 1990 to 2003. Am J Psychiatry 2005 Nov;162(11):2022-30

[17] Farlow MR, Graham SM, Alva G. Memantine for the treatment of Alzheimer's disease: tolerability and safety data from clinical trials. Drug Saf 2008;31(7):577-85

[18] European Public Assessment Report (EPAR) www.ema.europa.eu/htms/human/epar/a.htm

[19] NOTE FOR GUIDANCE ON GOOD CLINICAL PRACTICE (CPMP/ICH/135/95)

[20] Directive 91/507/EEC of the EU

[21] WMA Declaration of Helsinki - Ethical Principles for Medical Research Involving Human Subjects www.wma.net/en/30publications/10policies/b3/index.html

[22] Guideline on Medicinal Products in the Treatment of Alzheimer's Disease (CPMP/EWP/553/95) later replaced by GUIDELINE ON MEDICINAL PRODUCTS FOR THE TREATMENT OF ALZHEIMER'S DISEASE AND OTHER DEMENTIAS www.ema.europa.eu

[23] Grossberg, Pejovic, Miller and Graham (Dementia and Geriatrics 2009 27 164-172)

[24] Altman DG, Schulz KF, Moher D, Egger M, Davidoff F, Elbourne D, Gotzsche PC, Lang T: The revised CONSORT statement for reporting randomized trials: explanation and elaboration. Ann Intern Med 2001; 134: 663–694.

[25] Moher D, Schulz KF, Altman DG: The CONSORT statement: revised recommendations for improving the quality of reports of parallel- group randomized trials. Ann Intern Med 2001; 134: 657–662.

[26] Schulz KF, Altman DG, Moher D: CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials. BMJ 2010; 340: 698-702

[27] Molnar FJ, Man-Son-Hing M, Hutton B, Fergusson DA.Have last-observation-carried-forward analyses caused us to favour more toxic dementia therapies over less toxic alternatives? A systematic review. Open Med. 2009 Mar 24;3(2):e31-50.

[28] Livingston G, Katona C, Roch B, et al. A dependency model for patients with Alzheimer's disease: its validation and relationship to the costs of care--the LASER-AD Study. Curr Med Res Opin 2004 Jul;20(7):1007-16

[29] Sapin C, Livingston G, Katona C et al. Main factors influencing preference-based utility values in Alzheimer's disease: Results from the LASER-AD Study. (2005). Poster presented at the 7th International Conference on Alzheimer's and Parkinson's Disease (AD/PD), Sorrento, Italy, March 9-13

[30] Stern Y, Tang MX, Albert MS, et al. Predicting time to nursing home care and death in individuals with Alzheimer disease. JAMA 1997, 277:806-12

[31] Scarmeas N, Brandt J, Albert M, et al. Delusions and hallucinations are associated with worse outcome in Alzheimer disease. Arch Neurol 2005, 62:1601-8

[32] Scarmeas N, Brandt J, Blacker D, et al. Disruptive behaviour as a predictor in Alzheimer's disease. Arch Neurol 2007, 64:1755-61

[33] Guideline on medicinal products for the treatment of alzheimer's disease and other dementias. Committee for Medicinal Products for Human Use (CHMP), 2008. London: doc. ref. cpmp/ewp/553/595 rev. 551

[34] Vellas B, Coley N, Andrieu S. Disease modifying trials in Alzheimer's Disease: Perspectives for the future. J Alzheimer's Dis 2008, 15:289-301