Donepezil, galantamine, rivastigmine and memantine for the treatment of Alzheimer's disease (Review of TA 111)

Eisai/Pfizer Response to Technology Assessment Report (TAR)

4th August 2010

Eisai and Pfizer welcome the opportunity to comment on the Technology Assessment Report from the Peninsular Technology Assessment Group (PenTAG) for the review of TA111. In summary the key points we would like to highlight to the Appraisal Committee for consideration at the meeting on 25th August fall into three categories.

1) **The requirement for a sophisticated & sensitive approach in this complex disease area**
2) **Concerns regarding the overly-simplistic approach and lack of validity of the PenTAG model and results.**
3) **The robustness of the Eisai/Pfizer model, which shows cost savings in all plausible scenarios**

Modelling approach

- It is essential to choose a model structure which appropriately reflects the progression of Alzheimer's Disease (AD) and the potential benefits of treatment in order to accurately inform decision making in this important disease area. Indeed, in the 2004 appraisal the economic modelling was heavily criticised for a number of reasons such as excluding behavioural symptoms and a lack of sensitivity to changes in utility and costs outside of institutionalisation.

- The modelling approach chosen by PenTAG in this appraisal represents a step backwards in the development of an appropriate cost effectiveness framework on which to base recommendations. They detail that they have adopted a simplistic approach due to limitations of time, a lack of training in more advanced methods and unfounded assumptions about the lack of availability of individual patient level data.

- Eisai/Pfizer have developed a discrete event simulation model which:
  - Simultaneously predicts changes in cognition, function and behavioural symptoms over time
  - Has greater precision as it simulates the progression of Alzheimer's disease and treatment effects at the level of the individual
  - Allows for a more precise quantification of treatment effects over the entire course of the disease
  - Incorporates efficacy data for up to 12 months from clinical trials
  - Is sufficiently flexible to be adapted to include the other cholinesterase inhibitors and memantine if required.

The PenTAG model

- PenTAG have adopted a simple Markov–like model approach which has resulted in a flawed analysis which cannot be relied upon for decision-making. We

therefore recommend that this analysis should not be used by the Committee as a basis for its guidance. The key issues with the PenTAG model are as follows:

Equations underpinning model are invalid & predictions implausible

- The central equation predicting institutionalisation is based on a small (n=92) unrepresentative study population in Oxford. Moreover, the equation appears to be mis-specified and shows an extremely poor fit with the observed data from the study. As a result the predictions from the equation are unrealistic. For example, the equation predicts that a moderate disease patient who experiences an improvement in MMSE from 17 to 30 (implying full health) is associated with a median delay to institutionalisation of just over one month (36 days). Implemented in the model, the mean delay is estimated at 47 days. This is inconsistent with 6 months trial data showing cholinesterase inhibitors are associated with MMSE improvements of 1.37 and other data demonstrating donepezil is associated with delays in institutionalisation of up to 17.5 months.

- Equations used to predict costs do not take account of several outlier patients, nor control for patient characteristics, and the functional forms selected do not seem to fit with the observed data. As a result the predictions of the equations do not fit expectations or the data. For example, costs in the immediate time before institutionalisation are greater than those associated with institutionalisation, which is inconsistent with studies such as the Dementia UK report and previous modelling in this area.

- The probabilistic sensitivity analyses reveal the inappropriateness of the equations, as they indicate that improvements in cognition and function have a high probability of being harmful (i.e., leading to worse Quality Adjusted Life Years (QALYs)). This occurs in 26-40% of model runs in the base probabilistic sensitivity analyses. Eliminating these implausible runs from the analyses deliver donepezil costs/QALY of between £11,000 and £29,000.

- The Oxford study included the Barthel rather than the ADCS-ADL instrument to evaluate functional outcomes. The Barthel instrument is an older scale, not commonly used in AD and was developed to evaluate functioning in other diseases such as stroke. In the PenTAG model a mapping between the instruments was used. However, the mapping was based on 3 data points only, which do not contain over 60% of the range in possible ADCS-ADL scores. Reasonable alternative mapping equations produce dramatically different results from the model. For example, assuming a constant relationship between the two scales reduces the cost per QALY for donepezil by approximately £60,000 to £23,240.

Model does not reflect costs and QALY changes associated with AD

- The underlying prediction of the model is that, on average, patients who remain outside of institutional care, experience no change in cognitive function, costs or quality of life for approximately 15 years. When references to empty cells are corrected, MMSE, costs and QALYs do not change over the entire lifetime of the model. This is inconsistent with the studies used to populate the cost and quality of life data in the model and significantly underestimates the benefits of treatment while patients remain in the pre-institutionalised state.

Structure of model is incomplete as it excludes behavioural symptoms

- Whilst cognition and functional outcomes are incorporated, the clinical benefit of donepezil is still under-estimated in the model as the impact on behavioural symptoms is not incorporated. Pooled trial evidence from a systematic review demonstrates that donepezil has a significant impact on behavioural symptoms, resulting in a mean difference from placebo in NPI score of -1.76 (95% CI: -3.37; -0.15). This omission represents a step backwards in the progress made in modelling Alzheimer's disease in the previous appraisal, as developments such as the inclusion of behavioural symptoms have not been retained.

Key efficacy data excluded from base case cost effectiveness estimates
- 12 months efficacy data for donepezil are incorporated into the model, but are not used in the base case cost effectiveness analysis for donepezil. This approach biases the assessment against donepezil which has important and robust trial data beyond 6 months.

Model contains errors
- The PenTAG model contains technical errors such as referencing empty cells, allowing improvements in cognition and function to increase the risk of institutionalisation, and allowing negative numbers of patients to enter institutional care.

Eisai/Pfizer model outcomes

- The Eisai/Pfizer model demonstrates that donepezil is a cost effective therapy when mild and moderate AD patient sub-groups are considered as separate entities or as a combined group. The model finds donepezil is less expensive and delivers more QALYs than best supportive care.
  - Importantly, donepezil remains cost effective ('dominant') when key model parameters are varied and tested within reasonable bounds by the Decision Support Unit (DSU), and when errors noted were rectified.
  - The only exploratory analysis undertaken by the DSU where the cost effectiveness of donepezil becomes borderline is where the probability of institutionalisation is assumed to be independent of disease severity. This assumption is implausible and inconsistent with the literature and data PenTAG use themselves in their model. Further detail on our responses to the PenTAG criticisms are contained within this document.

Recommendation
- Overall, we believe the PenTAG model is not an accurate representation of the progression of Alzheimer's disease and does not capture the benefits of treatment. Indeed, PenTAG themselves state that their model is no more reliable than others submitted for this review. In light of this statement we would advocate that the Eisai/Pfizer model is a more appropriate model on which to base recommendations.

Responses to Critiques of the Eisai/Pfizer model

Eisai/Pfizer Model Results Robust to Alternative Analyses run by DSU/PenTAG

- The assessment group repeatedly state that running the model submitted by Eisai / Pfizer under 'alternative plausible' assumptions results in cost-effectiveness ratios that are at the border of what is normally considered cost-effective.

Response

- Although a number of changes were made to the model and input data, only one change resulted in a scenario where donepezil did not dominate best supportive care (i.e. show costs savings versus best supportive care).
- This change was where the DSU modified the model inputs to set the proportion of patients institutionalised as equal across all disease severity levels.
- This change assumes institutionalisation is completely independent of disease severity, which is not only in direct contrast to the premise of the PenTAG model, but also contradicts all available evidence (Wolstenholme et al. 2002; Wattmo et al. 2010; Haupt & Kurz 1993; Gaugler et al. 2009). We contend that this change is not clinically plausible and alternative more plausible estimates would have produced highly favourable cost per QALYs.
- Thus the Eisai/Pfizer model is robust to all reasonable and plausible analyses run by the assessment group and so would have fulfilled the requirements of a typical single technology appraisal (STA).

CERAD dataset is representative of natural progression of cognitive decline in Alzheimer's disease in UK

- PenTAG assert that " . . .the participants in the US CERAD study are not described in any detail, thus it is unclear how representative they are of UK individuals with mild to moderate AD. For example, it is stated on page 89 of the main submission that the CERAD data base does not include 'treated' patients. Little further discussion of this point is provided but it suggests that individuals included in the study might not necessarily be representative of a typical mild to moderate AD population" (section 6.3.4.1)

Response

- The CERAD data are not used to define characteristics of the UK population with Alzheimer's disease, but only to estimate changes in MMSE over time in untreated patients. There is no reason to believe that the underlying progression of MMSE in a US population would be substantially different than that in a UK population, and we feel the large sample (721 patients) leads to greater accuracy than basing estimates on a UK database with a much smaller sample (e.g., 92 patients in the study used to inform the PenTAG model). The CERAD registry is also one of the best known Alzheimer's disease registries in the field and there are extensive publications describing the population in the registry (Fillenbaum GG 1997, Heyman 1996, Heyman 1997, Morris JC 1993, Neumann 2001). Moreover, Appendix H in the Eisai/Pfizer submission provides a description of the cohort used to develop the MMSE equations. As a point of clarification, participants in the CERAD registry were not treated with cholinesterase inhibitors, as the timeframe of the study pre-dated use of cholinesterase inhibitors. We also note that sensitivity analyses around the background rate of disease progression were run in the Eisai/Pfizer submission.

Goodness of fit of disease progression equations is assessed

- The TAR states "Additionally, corresponding model statistics, such as goodness of fit, are not provided and there has been no to attempt to validate the MMSE risk equation against external data sources, a point noted by the authors of the original economic model " (6.3.4.1)

Response

- Appendices H and I in the Eisai/Pfizer submission provide additional information on derivation of the equations, and Figure 58 in the PenTAG report shows the predicted versus observed changes in MMSE from CERAD produce a very good fit. Standard errors are reported for all coefficient estimates. We acknowledge that external data sources were not obtained to conduct external validation, but the uncertainty around these parameter estimates was explored through sensitivity analyses. We note that the PenTAG equations were also not validated against external sources.

CERAD data is appropriate primary source for disease progression

- PenTAG state "the manufacturer notes that the annual rate of change in MMSE was notably different when RCT data were used instead of individuals from the CERAD study; this point is to some extent illustrated in Figure 58 (Figure 4 in Appendix H of the submission). Specifically the submission states that using the alternative source of data led to 'no change or a small annual change in MMSE scores <20 and potentially large declines for those with values above 20'. A reason for this possible discrepancy is suggested - shorter measurement intervals in controlled studies - but it is uncertain that this in itself is sufficient justification for choosing one source over another, or whether it indeed suggests more reason to use it as the primary source."

Response

- The TAR comment is not a complete description of the rationale for the decision to use CERAD rather than the trial data. As indicated in Appendix I of the Eisai/Pfizer submission, "While MMSE data over time were available from trial data, the CERAD data offered a longer time course of data. Furthermore, the patterns of change observed in CERAD were more in line with what has been previously reported on progression of Alzheimer's disease, with progression slowest over the mildest and most severe stages of the disease. Figure 2 (**Figure 58 in the Assessment group's report**) plots average annual MMSE rates of change on the Y-axis by previous MMSE score on the X-axis. Results are plotted for the observed CERAD data and for predicted CERAD data using the CERAD based equation. Observed changes are also plotted for treated and untreated patients in the donepezil clinical trials. As Figure 2 indicates, the trial data showed a positive annual rate of change in MMSE (i.e., improvement) over some ranges of MMSE, even in untreated patients. Using the trial data to model the natural history of MMSE changes in an untreated population would not have been appropriate, as it would have led to predictions of improved cognition in a subgroup of untreated patients."

Additional information on ADL/IADL transformation

- In section 6.3.4.2 the assessment group implies that the only description of the ADL/IADL transformations was "trials measuring ADL and IADL used a variety of scales so 'standardised scales' were constructed using items from the various measures in order to link trial results to the utility function".

Response

- Appendix H of the Eisai/Pfizer submission provides a detailed description of the transformation and development of the equations. The approach is very similar to that adopted by Gauthier at al 2010[*]. It is worth noting that in the submitted analyses, patient ADL and IADL scores only influence caregiver QALY estimates. If the model is run without any benefit of treatment for ADL and IADLs, donepezil is still predicted to lead to improved patient QALYs and lower costs compared to best supportive care.

Effect of donepezil on NPI, ADL and IADL not double counted

- The DSU contends that treatment effect for NPI, ADL, and IADL may be overestimated since the equations include a term for treatment, but also terms for current MMSE scores (see section 6.3.4.3). Since donepezil increases MMSE scores, they feel that the treatment effect coefficient may be overestimated.

Response

- As described in the technical appendices to the Eisai/Pfizer submission, there is no double counting in the equations. Treatment may affect NPI, ADL and IADL through different mechanisms, including possibly a delay of cognitive decline. That is, part of the improvement seen in these measures could be in part the result of the treatment's effect on MMSE. Treatment may also affect NPI, ADL and IADL over and above its effect on cognition. Including current MMSE in the equations captures these two mechanisms, without double-counting the effect. The coefficient for treatment is reduced by the inclusion of current MMSE as it now represents the effect of treatment over and above changes in cognition. The total treatment effect is still maintained in the equation and reflected in predictions for treated and untreated patients. These are reported in Appendix J which shows that the equations provide very good estimates of treatment effect sizes, with the exception of IADL, where treatment effects seem to be underestimated.
- Including current MMSE is important since it provides greater predictive ability, and more importantly, it captures the natural correlation that will exist between all of the measures over time. Thus, changes in a given direction for MMSE will lead to consistent changes in the other measures.

Choice of cutpoint for donepezil effect in cognition disease progression equation

- PenTAG question the basis of the assumption that the treatment of donepezil in the cognition disease progression equation is different after 20 weeks (see section 6.3.4.3).

Response

- The choice of a cut point at 20 weeks for estimating treatment effect was based on visual evaluation of the data, which indicated a marked difference in rate of MMSE change between treated and untreated patients over the first 20 weeks, but a much smaller, though still significant, difference thereafter. These results are consistent with the 6 and 12-month meta-analysis results on MMSE for donepezil reported by PenTAG in the model, which indicate that the treatment effect size after 12 months for MMSE (1.720) is greater than after 6 months (1.237), but that the majority of the treatment effect is observed over the first half of the year.
- Appendix A provides a graphical depiction of the data used to inform the 20 week cut-off point.

Effect of donepezil on patient utility not double counted

---

[*] Some studies included severe AD patient populations (out of licence for donepezil)

- Concern is expressed by PenTAG that including both MMSE and NPI in patient utility equations is double counting the effect of treatment (see section 6.3.4.4).

Response

- As noted above, the treatment effect for NPI is not being double counted, as the treatment effect term for NPI is calculated after controlling for changes in MMSE.

Transformation of the NPI coefficient for utility calculations

- The DSU notes that the transformation for the NPI coefficient in the patient utility calculation is not described (see section 6.3.4.4).

Response

- The source paper for the utility equations used a version of the NPI which had a range of scores of 0 to 30, whereas the trial NPI data used the traditional 0 to 144 scale. The coefficient for NPI in the utility equation was therefore reduced by multiplying it by 30/144. Had such a transformation not been applied, the model would have overestimated the effect of improvements in patient NPI scores on health utilities.

Caregiver proxy responses more appropriate than patient responses to model patient utilities

- PenTAG criticise the Eisai/Pfizer approach of using caregiver proxy responses rather than patient responses (see section 6.3.4.4).

Response

- However, as noted by PenTAG themselves (see page 293), caregiver proxy utilities are felt to be more appropriate in this population, as patients themselves may not provide reliable responses.
- As Table 85 in the DSU's critique indicates, using the patient responses would have led to implausible estimates. For example, patients with MMSE scores under 10, had better utilities than those with MMSE scores of 10-15, and indeed health utilities that would be comparable to individuals of a similar age without dementia.

Caregiver utilities are unreliable but were extensively tested

- PenTAG express various concerns about the equation which estimates caregiver utilities (see section 6.3.4.5).

Response

- The caregiver utility equation was estimated using data from those donepezil clinical trials in which caregiver quality of life was reported using the SF-36. The explanatory variables included in the equation are those which were found to be significant based on our analyses, as described in Appendix J of the Eisai/Pfizer submission. While we must reiterate that treatment effects are not being double counted, we acknowledge that the equation used to predict caregiver utilities does not represent a definitive estimate and is subject to considerable uncertainty (see section 6.3.4.5). We tested caregiver utility inputs in sensitivity analyses and even with the exclusion of an effect altogether, the direction of effect remains unchanged.

Rates of institutionalization are appropriate

- The DSU provides a number of minor critiques of institutionalisation rates used in the model (see section 6.3.4.6).

Response

- In terms of the institutionalisation rates, two data sources were used for this prediction: a study of institutionalised patients that showed that 67% of the population with dementia had MMSE scores indicative of moderate to severe disease, and 33% with MMSE scores indicative of mild disease (Macdonald 2007); and the Dementia UK report (Knapp 2007) which reported that 36.5% of the entire population with dementia were living in institutional care, and that 55% had moderate to severe disease and 45% had mild disease. Simple algebra can then be used to come up with an approximation of the percent of patients in institutional care by disease severity. If the UK Dementia figures are correct, and the Macdonald data, which were used by the study authors to project the number of dementia cases in long term care in the United Kingdom, are representative of disease severity in institutional care in the UK, then it should follow that: 45%*X + 55% *Y = 36.5%, where X = the percent of patients with mild disease in institutional care, and Y, the percent with moderate to severe disease in institutional care. Further, it follows that 45%*X / 55%*Y = 67%/33%. By solving for X and Y, we can come up with an estimate that 54.8% of individuals with moderate to severe disease reside in institutional settings, compared with 21.8% of individuals with mild disease. If we then use the mid-points of MMSE scores representative of mild disease and moderate-severe disease (as used in the Eisai/Pfizer model), we can come up with proportions across different severity levels. While we acknowledge that these calculations are subject to considerable uncertainty, they are in line with those cited by PenTAG from the LASER-AD study, and in fact produce less differentiation in rates of institutionalisation by disease severity than those from LASER-AD. We also note that the findings in Macdonald 2007 are comparable to an earlier study conducted in the UK (Darton 1998).
- The graph below shows institutionalisation rates (y-axis) at different levels of MMSE (x-axis) used in the Eisai/Pfizer submission (DES), those from LASER-AD, extrapolated across all ranges of MMSE (LASER), and the 'plausible' alternative assumption used by the DSU.
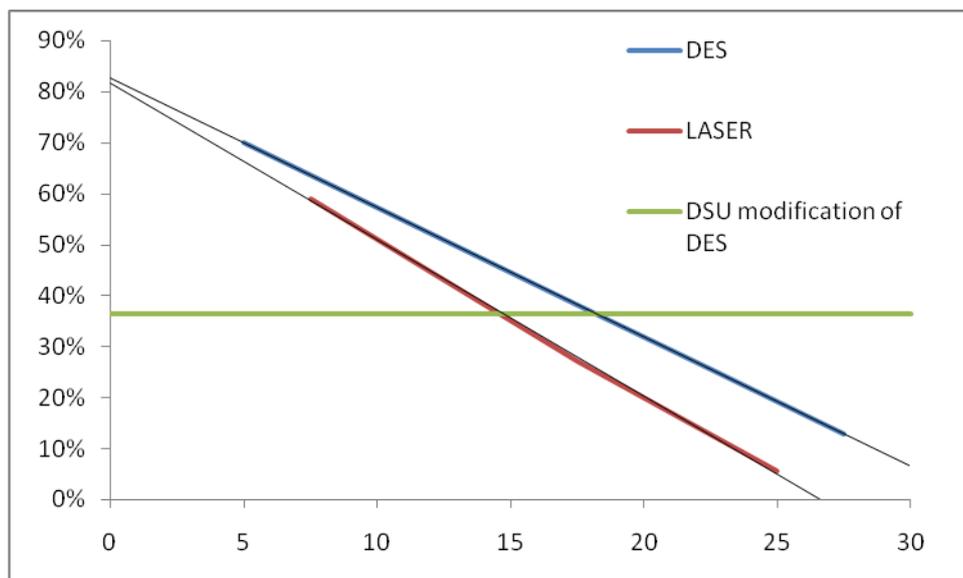
Cost estimates are appropriate

- Overall, the DSU provides a number of minor critiques of cost estimates used in the model (see section 6.3.4.6), but we would like to highlight that costs were tested extensively in sensitivity analyses.
- The cost used to assign monitoring physician visits for patients on donepezil is based on: "Service Code: 430, Geriatric Medicine. National Average Unit Cost £60.65 (Follow-up, non-admitted, face to face)" from the 2007-2008 NHS combined file. This value was inflated to 2009 currency levels for a final estimate of £62.29.
- The cost used in the Eisai/Pfizer model for donepezil 10mg of £3.00 is correct. The price of all doses and formulations of donepezil was reduced by 5.8% under the PPRS agreement in January 2010.
- Costs of care for patients residing in the community were taken from the Dementia UK report and assigned based on MMSE ranges based on the assumption that the costs reported in Dementia UK for mild Alzheimer's disease could be assigned to patients with MMSE scores at or above 20, costs for moderate Alzheimer's disease to patients with MMSE scores between 10 and 19, and costs for severe Alzheimer's disease to patients with MMSE scores below 10.
- Whilst we agree that the costs data used from the Dementia UK report are based on a small sample (114 individuals) and maybe out of date, we note that the cost estimates are based on a larger sample than those used in the PenTAG model. The PenTAG model also used estimates from the late 1990s.

Data availability limits ability to replicate UK population

- PenTAG raise a technical concern with the patient population modelled in the Eisai/Pfizer submission, namely that the model samples from a population that produces a population that is marginally younger for moderate disease than for mild disease (see section 6.3.5.1).

Response

- This was true in the trial data, and true after the age and sex sampling weights were adjusted in the model to replicate the age and sex distribution of the UK population with Alzheimer's disease.
- We did not have UK-only patient level data to reflect directly the severity of disease by age and sex to be representative of the UK population. So while the simulated population may not be completely reflective of the UK Alzheimer's disease population, this is not due to technical errors in the model, but a consequence of the available data. As the DSU analyses indicate, changing survival inputs in the model does not substantively alter results, since survival is assumed to be independent of treatment.

Model calculates rate of change in MMSE rather than annual increments

- The DSU reports that MMSE is calculated based on annual increments (see section 6.3.3.2 & 6.3.4.1).

Response

- This is not the case. The model calculates the rate of change in MMSE as a function of time. The equation used in the model is parameterised in terms of annual rate of change since some time parameterisation was necessary, but an alternate parameterisation (e.g., daily) would have produced equivalent results. MMSE change in the model is time dependent, as are treatment effects on MMSE.

Sensitivity analyses presented in Eisai/Pfizer model were extensive

- While we accept that there are number of uncertainties in the model inputs, the DSU assessment does not give a balanced picture of the extent of sensitivity analyses presented in the Eisai/Pfizer submission, which included reducing the cost of institutional care by 30%, reducing institutionalisation rates by as much as 50%, reducing treatment effects by 50%, etc. Many of the questions the DSU had on the potential consequences of changes in these inputs are addressed by these analyses.

Modifications made by DSU are appropriate but do not change the results substantively

- We agree with the modification made by the DSU to the time to discontinuation calculations (see section 6.3.5.3) and note that net costs with treatment changes by <£10/patient and net QALYs by 0.001 per patient.
- We agree with changes made by the DSU to correct technical errors identified in the model (see sections 6.3.6 and 6.3.7), but note that the net effect of these changes improved outcomes for donepezil versus supportive care, including the results of the probabilistic sensitivity analyses.
- Changes to MMSE scaling, life expectancy, hazard calculations, and the beta distribution parameters for institutional care made by the DSU were appropriate (see section 6.3.8).
- We note that the net effect of the changes made by the DSU were to make results somewhat more favourable for donepezil relative to best supportive care.

Eisai/Pfizer model robust to plausible exploratory analyses undertaken by DSU

- As mentioned, we do not agree that assuming that the proportion of patients institutionalised is equal across all levels of disease severity represents a plausible scenario (see section 6.3.9).
- We believe that removing the coefficients for NPI, ADL, and IADL from the caregiver utility equations once a patient reaches institutional care, is an unduly harsh assumption, leading to the curious prediction that a treatment that delays disease progression is harmful to caregivers. We note that the adjustment by the DSU could lead to improvements in some individual caregiver utilities of as much as 0.22 (on a scale of 0 to 1) once patients enter institutional care. We also note that even with this harsh assumption, donepezil is still dominant over best supportive care.
- Although we have noted that we do not feel that the model overestimates treatment effects for NPI, ADL, and IADL, we note that the change made by the DSU, had little impact on results.

Modifications to the regular update interval produce expected changes in model predictions

- The DSU questions the reliability of the model because it cannot explain why results for donepezil improve when the regular update interval is reduced.
Response
- Reducing the update interval leads to more accurate depictions of disease progression and treatment effects, and it is expected that a shorter interval would lead to better results for donepezil because it more finely captures the delay in progression from one level of disease severity to another, and as a consequence, the effect on costs and utilities. If for example, the update interval was set to 10 years patients would get processed for the first time after entering the model at 10 years, by which time the vast majority, if not all patients, will be in the very most severe

stages of the disease regardless of whether they receive donepezil or not. In this scenario patients would not receive any benefits of treatment as they had not been processed prior to the 10 year time point, but would incur drug costs. More frequent update intervals allows more events to happen and disease progression to be captured in a more accurate and detailed way. Whilst daily update intervals would be the most accurate, this increases the processing time of the model and so we opted for a 90 day interval as a compromise. As the DSU notes, no half-cycle correction was included in the model (see section 6.3.5.2). It was felt with update intervals set to a maximum of 90 days, this was unnecessary.

Additional clarifications on issues raised by the DSU and PenTAG on the Eisai/Pfizer model are presented in Appendix A.

Eisai/Pfizer Comments on the PenTAG Approach

- The decision by PenTAG to adapt a simplified modelling technique despite the well-acknowledged limitations associated with oversimplification of modelling in this area is highly questionable. Review of the PenTAG model identified a number of concerns, but we focus on the critical flaws in the PenTAG model and consequent analyses. Additional concerns about the model and inputs are provided in the Appendices B and C, and in the Pro-forma responses based on review of the executable model.

Central equation predicting institutionalisation is flawed

- One of the critical flaws in the PenTAG model is that the equations that predict time to institutionalisation lack face validity and are technically flawed.
    - The finding that neither cognition nor function are significant predictors of institutionalisation runs contrary to an extensive body of research which suggests the opposite (Wolstenholme et al. 2002; Wattmo et al. 2010; Haupt & Kurz 1993; Gaugler et al. 2009). The finding by PenTAG is especially surprising in that a previous study using the same dataset and published by Wolstenholme and colleagues found that both MMSE and Barthel score were independent significant predictors of time to institutionalisation (Wolstenholme et al. 2002).
    - Appendix C describes in more detail how the model fit to predict time to institutionalisation does not accord with the observed data. The model used to predict risk of death also does not fit the observed data, as outlined in the Appendix C.
    - A decision was made to include MMSE and ADL in the model even though these were not statistically significant, and the coefficients on the MMSE and ADL terms highly unreliable. The unreliability of the equations is highlighted in the probabilistic sensitivity analyses, which suggest that there is a strong possibility that treatment is harmful (i.e., reduces QALYs). Since there are no disutilities in the model associated with adverse events, and variation in the treatment effect sizes do not lead to scenarios where treatment is less effective than best supportive care, the only possibility for this finding is that the time to institutionalisation and time to death equations predict that in a substantial number of cases (26% in base case), a treatment that improves MMSE and ADL will lead to higher rates of institutionalisation. If one excludes the 26% of model runs where this occurs, the mean ICER for donepezil predicted by the probabilistic sensitivity analyses drops to £28,385. If one excludes all model runs where either improvements in MMSE are predicted to be harmful (40% of runs in the base case) or improvements in the Barthel Index are predicted to be harmful (26% of runs in the base case), then the mean ICER drops to £11,865.
    - Of note, a difference of one standard error in the coefficient for MMSE in the equations would lead to results where donepezil was dominant over best supportive care (see Pro-forma responses based on review of executable model).
    - It is worth noting that when the time to institutionalisation equation is implemented outside of the model, the median delays to institutionalisation are small even for very effective treatments. For example, for a population with a mean age of 77, mean MMSE of 17, and mean ADL score of 17.52, an increase in MMSE of 6 points

12

results in a difference in median time to institutionalisation of 16 days (if no individual dies before being institutionalised). An improvement in MMSE to 30, results in a difference of 36 days. Actual model predictions are provided in the Pro-forma responses based on review of the executable model. This is inconsistent with 6 months trial data showing cholinesterase inhibitors are associated with MMSE improvements of 1.37 (Birks 2006) and other data demonstrating donepezil is associated with delays in institutionalisation of up to 17.5 months (Geldmacher 2003).

- While seeking a UK data source to establish a UK-specific risk for institutionalisation is appropriate, the data source used by PenTAG is problematic and the methods are too flawed to have any reasonable expectation that the equations developed would be reliable. There are a number of problems:

  - The sample size is small: 92 patients in total, and only 22 with mild Alzheimer's disease at study entry.

  - Although MMSE and ADL scores were available for individuals over the course of the study, PenTAG opted only to use baseline MMSE and ADL as potential predictors. A more appropriate approach would have evaluated how changes in MMSE/ADL, and 'current' MMSE/ADL scores influence the risk of institutionalisation.

  - As far as we can tell, the equation only considered three potential predictors of institutionalisation and death: baseline MMSE, baseline ADL and age, and only one functional form (one that assumes that the risk of institutionalisation is constant over time). No goodness of fit statistics are provided, and there is no indication that alternative models were tested, although it is clear from Figure 64 that the predicted values do not match the observed data

  - By using a prevalent population residing in the community, there are inherent biases in trying to use severity at study entry only to predict outcomes like institutionalisation. Patients with severe disease in the sample had managed to remain in the community, despite having very low MMSE scores (average 5.4) and ADL scores (average 14.92). These patients are more likely to have other characteristics which predispose them to remaining in the community compared to patients entering the study with much more mild disease (and likely a much shorter duration of disease). This bias could greatly diminish any relationship between disease severity and institutionalisation.

  - Given the small sample size, one would have expected that considerable attention would be paid to the quality of the data, but Figure 66 in the PenTAG report, which plots MMSE scores by time to institutionalisation reveals some puzzling information. For example, one individual's MMSE score appears to drop below zero (the scale runs from 0 to 30), and another's appears to have remained at 0 for at least 2 years before that individual was institutionalised. The degree of variation in these figures also highlights how unstable estimates based on these data will be given the small sample size, as there is no indication from the PenTAG report that any effort was made to account and correct for outliers that may have had a disproportionate influence on outcomes.

- Finally, we note that there is a logical inconsistency in the equations. PenTAG developed equations which allow for the prediction of MMSE by time to institutionalisation. For mild to moderate disease, these equations indicate that a 1 year difference in time to institutionalisation is associated with a 4.17 difference in MMSE. The same equation (MMSE = 8.34 + 4.17t) can be used to solve for time to institutionalisation for a given MMSE score. A 1 point difference in MMSE would result in an almost 3 month delay in institutional care. This is inconsistent with the time to institutionalisation equation, which predicts much smaller delays. Not only is this inconsistent, but it leads to an incoherent model. A 1 point change in MMSE will lead to roughly a 10 day delay in institutional care according to the PenTAG equation. This 10 day delay is then used to calculate differences in MMSE scores in the pre-institutional care state. The 10 day delay, however, means that the original 1 point difference in MMSE is transformed to a difference equal to 4.17*10/365, or 0.11.

Equations used to predict utilities and costs are flawed
- As mentioned previously, the equation used to predict MMSE over the course of pre-institutionalisation is incoherent with estimates in time to institutionalisation. Because the effect of MMSE on time to institutionalisation is so small in the PenTAG model, the MMSE difference attributed to patients before they become institutionalised is also much smaller than the MMSE treatment effect entered into the model. Since the assignment of utilities is based on these MMSE calculations, this severely underestimates the potential benefits of treatment over this period. Other problems with the MMSE and costs equations include the following:
    o Since not all of the patients in the dataset analysed by PenTAG were institutionalised, it is unclear how data for patients who were never institutionalised can factor into calculations predicting costs as a function of time to institutionalisation. MMSE is also calculated based on time to institutionalisation so suffers from the same problem.
    o We note that the equations developed by PenTAG to predict MMSE in mild to moderate disease do not allow individuals' MMSE scores to drop below 8.34, and do allow them to exceed 30 (Note: in the model, this is corrected by artificially forcing the maximum MMSE score to be 27).
    o The study by Wolstenholme and colleagues (using the same data source used by PenTAG) (Wolstenholme 2002) found that ADL was a much stronger predictor of costs prior to institutionalisation than MMSE, yet PenTAG opted to build the cost relationship based on MMSE.
    o The function developed to relate costs to time to institutionalisation does not control for any other individual characteristics, and as evidenced by Figure 70, does not fit the data very well (See Appendix C). There is no indication that PenTAG made any attempt to deal with outliers, deal with non-normally distributed costs or tested alternative functional forms beyond adding polynomial terms to the predictive equation.
    o We also note that the cost equations used for mild to moderate disease will predict negative cost for some patients, and as indicated by Figure 70 (note: in the model PenTAG corrects this by artificially assigning a maximum time to institutionalization

of 75.5 months), for mild to moderate patients, predict costs lower than those observed in the dataset for individuals whose time to institutionalisation exceeds roughly four years.

- o The equations also predict that for a substantial period of time, costs of care in the community are higher than those in institutional care. This runs contrary to what has been reported in the Dementia UK report, which indicated that even for severe disease, costs of care in the community are considerably lower than costs of care in institutions.

- o A review of the model indicates **that cost and utilities for patients who are not institutionalised remain constant for upwards of 15 years** (See Appendix B). Further, the changes in these outcomes after 15 years are a result of an error in the model, whereby empty cells are referenced as part of the calculations. Thus although PenTAG claims to allow for a gradual increase in costs and decrease in utilities while patients remain in pre-institutional care, this does not happen for most of the modelling period, and only begins once the great majority of patients have either died or entered institutional care. Underlying these estimates, one can observe that MMSE scores in the model base case do not change in approximately 15 years in the cohort remaining outside of institutional care. When the referencing of empty cells is rectified MMSE scores, costs and QALYs do not change for the entire lifetime of the model.

Important elements of donepezil efficacy excluded by PenTAG

- Although the review by PenTAG identified behavioural symptoms as an important component of Alzheimer's disease in terms of predicting quality of life and costs, the model that was developed does not consider behavioural symptoms at all.
- The importance of considering behavioural symptoms was acknowledged in the previous 2004 appraisal, was included in the scope for this review and there is evidence that donepezil has a statistically significant impact on NPI (Standardised Mean Difference -1.76, 95% CI, -3.37, -0.15) (Campbell et al., 2008[†]).
- An analysis that is not capable of evaluating the impact of this component of Alzheimer's disease is incomplete and excludes potential benefits associated with treatment.

Mapping of ADL scores to Barthel Index uncertain

- The equations used to transform ADCS-ADL scores to Barthel scores are almost completely arbitrary and based on only 3 data points that exclude over 60% of the ADCS-ADL scales.
- The form selected by PenTAG required large changes in ADCS-ADL for patients with mild to moderate Alzheimer's disease, for any substantive changes in the Barthel Index. This would suggest that the two scales are not highly correlated and that the Barthel Index itself may not be a good indicator of function. This may in part be due to the fact it is an old scale originally developed for patients with stroke. Alternatively, the function form used by PenTAG may not have been appropriate and alternative forms produce very different results (see Pro-forma responses based on review of executable model). For example, a linear transformation reduces the ICER for donepezil by almost £60,000 to £23,240.
- The function used by PenTAG also results in some implausible predictions. For example, the function predicts that improvements in ADCS-ADL are associated with deterioration in

---

[†] Some studies included severe AD patient populations (out of licence for donepezil)

Barthel scores for ADCS-ADL scores above 75. An improvement in ADCS-ADL from 60 to 70 (**10 points**) is equivalent to a **+0.66** change in Barthel score. From 70 to 80, it is equivalent to a change of **-0.66**.

- The uncertainty around the transformation included in the probabilistic sensitivity analysis is inadequate as it assumes the functional form selected by PenTAG is correct, and uses standard errors on a curve fit to three average response data points.

PenTAG model contains serious technical errors

- The model suffers from several serious technical errors (See Appendix B and Pro-forma responses based on review of the executable model) including referencing empty cells for about 30% of the modelling time horizon.
- The probabilistic sensitivity analyses show that the model allows for scenarios where the model predicts negative numbers of patients in the institutional care state (and assigns negative costs and utilities as a consequence).
- This problem occurs because at present, the equations to predict institutionalisation and death are applied independently in the model, so no matter what mortality data are entered into the model, the number of patients in pre-institutional care remains unchanged. It is worth noting that in this respect, the model is not a true Markov model and the model diagram presented by PenTAG is not accurate as it indicates that patients can transition from pre-institutional care to death (they do not in the model). This causes errors in the probabilistic sensitivity analyses, but is also a structural flaw in the model since mortality in the population does not influence the proportion of the cohort alive and residing outside of institutional care.

PenTAG exclude donepezil 1-year efficacy data

- In section 4 PenTAG note that "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" (p. 34) . However, there are double-blind RCT data for donepezil vs. placebo that appear to have been overlooked in the current PenTAG clinical effectiveness analyses.
- The PenTAG donepezil vs. placebo meta-analyses used RCT evidence identified in the new review and the NICE 2004 review but only included data up to the 6-month time point.
- However, two randomised controlled trials, (Winblad et al., 2001; Mohs et al., 2001) included in the NICE 2004 review compared the effects of donepezil vs. placebo on cognitive and functional outcomes in AD patients over one year, so it is unclear why only 6-month data from these studies are included in the PenTAG assessment of clinical effectiveness of donepezil.
- Winblad 2001 (N=286) reported a statistically significant difference on MMSE score, global assessment , and activities of daily living favouring donepezil vs. placebo at 52 weeks (treatment difference in least squares mean change from baseline, based on ITT LOCF : MMSE 1.69, p <0.001; Gottfries-Brane-Steen scale -3.72, p<0.05; Progressive Deterioration Scale 4.03, p<0.05)
- Mohs 2001 (N=431) reported a statistically significant difference on MMSE and functional decline favouring donepezil vs. placebo at 54 weeks (adjusted mean change between treatments from baseline: MMSE (ITT) p<0.001; AD Functional Assessment and Change Scale (completers analysis) p<0.001).
- It is unclear why PenTAG have omitted these two methodologically sound studies that report one year data for donepezil on several outcomes relevant to the clinical effectiveness

analyses. Especially, as in the cost effectiveness model a 12 months pooled effect size is reported for donepezil. This analysis and results should be included in the TAR.

Inclusion of AD2000 trial is inappropriate

- We note once again that PenTAG have included the AD2000 study (Courtney, 2004) despite the following multiple methodological limitations associated with this study (as stated in our submission and by the wider research community, e.g. Birks 2006):
    - o The study was under-powered; it set out to recruit 2000 patients but only recruited 565. It therefore had insufficient patients to carry out the probability calculations necessary to determine the statistical significance of the trial's outcomes.
    - o Entry criteria were based upon an "uncertainty principle" which was biased in favour of patients who were unlikely to respond to donepezil.
    - o The study population was mixed (i.e. it included patients with non-AD dementia).
    - o Although not affecting the 6 months or 12 months data, the treatment design (i.e. multiple washout periods) does not reflect current medical practice. Indeed, data have shown that patients who stop treatment for a short period lose their initial treatment benefit (Doody 2001 (reported in original submission); Burns 2007 (reported in current submission)). The design of this study renders it unsuitable for pooling with other donepezil studies (see Birks 2006).
    - o Given the above limitations, it is unclear why PenTAG have included this study in the pair-wise meta-analyses reported in the clinical effectiveness section of the TAR.

References

Birks J. Cholinesterase inhibitors for Alzheimer's disease. *Cochrane Database Syst Rev.* 2006(1):CD005593.

Burns A, Gauthier S, Perdomo C. Efficacy and safety of donepezil over 3 years: An open-label, multicentre study in patients with Alzheimer's disease. *Int J Geriat Psychiat.* 2007;22(8):806-812.

Campbell N, Ayub A, Boustani MA, et al. Impact of cholinesterase inhibitors on behavioral and psychological symptoms of Alzheimer's disease: A meta-analysis. *Clin Interv Aging* 2008;3(4):719-728.

Courtney C, Farrell D, Gray R, et al. Long-term donepezil treatment in 565 patients with Alzheimer's disease (AD2000): Randomised double-blind trial. *Lancet.* 2004;363:2105-2115.

Darton R, PSSRU survey of residential and nursing home care. In Knapp M, ed. Mental Health Research Review. Canterbury, Kent: University of Kent 1998, 28:26-30).

Doody RS, Geldmacher DS, Gordon B, et al. Open-label, multicenter, phase 3 extension study of the safety and efficacy of donepezil in patients with Alzheimer disease. *Arch Neurol.* 2001;58(3):427-433.

Fillenbaum GG, Beekly D, Edland SD, et al. Consortium to Establish a Registry for Alzheimer's Disease (CERAD): Development, data base structure, and selected findings. *Top Health Inform Management* 1997;18:47-58.

Gaugler JE, Yu F, Krichbaum K, et al. Predictors of Nursing Home Admission for Persons with Dementia. *Med Care* 2009;47:191-198

Gauthier S, Lopez OL, Waldemar G, et al. Effects of donepezil on activities of daily living: integrated analysis of patient data from studies in mild, moderate and severe Alzheimer's disease. *Int Psychogeriatr* 2010 Jun 10:1-11. [Epub ahead of print]

Geldmacher DS, Provenzano G, McRaeT, et al. Donepezil is associated with delayed nursing home placement in patients with Alzheimer's disease. *J Am Geriatr Soc* 2003;51:937-944.

Haupt M, Kurz A. Predictors of nursing home placement in patients with Alzheimer's Disease. *Int J Geriatr Psychiatry* 1993;8:741-746.

Heyman A, Peterson B, Fillenbaum G, et al. CERAD Part XIV: demographic and clinical predictors of survival in patients with Alzheimer's disease. *Neurology* 1996;46:656-660.

Heyman A, Peterson B, Fillenbaum G, et al. Predictors of time to institutionalization of patients with Alzheimer's disease. The CERAD experience, Part XVII. *Neurology* 1997;48:1304-1309.

Knapp M, Prince M, Albanese E, et al. Dementia UK: The Full Report. London: Alzheimer's Society; 2007: http://www.alzheimers.org.uk/site/scripts/download_info.php?fileID=2. Accessed December 9, 2008

Macdonald A, Cooper B. Long-term care and dementia services: an impending crisis. *Age Ageing* 2007;36(1):16-22.

Mohs RC, Doody RS, Morris JC, et al. A 1-year, placebo-controlled preservation of function survival study of donepezil in AD patients. *Neurology* 2001;57(3):481-488.

Morris JC, Edland S, Clark CM, et al. Consortium to Establish a Registry for Alzheimer's Disease (CERAD) Part IV. Rates of cognitive change, a longitudinal assessment of probable Alzheimer's disease. *Neurology* 1993;43:2457-2465.

Neumann PJ, Araki SS, Arcelus A, et al. Measuring Alzheimer's disease progression with transition probabilities: estimates from CERAD. *Neurology* 2001;57:951-964.

Wattmo C, Wallin AK, Londos E, et al. Risk factors for nursing home placement in Alzheimer's disease: a longitudinal study of cognition, ADL, service utilization, and

cholinesterase inhibitor treatment. The Gerontologist 2010 June 20 [Epub ahead of print]

Winblad B, Engedal K, Soininen H, et al. A 1-year, randomized, placebo-controlled study of donepezil in patients with mild to moderate AD. *Neurology* 2001;57(3):489-495.

Wolstenholme J, Fenn P, Gray A, et al. Estimating the relationship between disease progression and cost of care in dementia. *Br J Psychiatry* 2002;181:36-42

Appendices

Appendix A: Clarifications on issues raised with Eisai/Pfizer model

Several clarifications need to be made in response to the DSU' and PenTAG's critique of the economic model submitted by Eisai /Pfizer.

1. We do not agree with the assessment that the model submitted is not a 'pure' discrete event simulation because some of the cost and utility assignments incorporated into the model use average values.   All models, discrete event simulation, or otherwise, rely to some extent on some level of aggregation. While the submitted model attempts to provide as fine an estimate as possible, the model is flexible enough to incorporate 'average' inputs (e.g., average costs by ranges of MMSE, or proportion institutionalised) where more precise data are not available.

2. Contrary to what is reported, the QALYs are not the only health outcome produced by the simulation. In addition to QALYs, the model predicts outcomes related to time spent in institutional care, and time with severe Alzheimer's disease, which can be based on either MMSE, ADL, IADL, or NPI outcomes. The model also calculates caregiver time outcomes, although no cost is attributed to caregiver time in the submitted analyses.

3. Contrary to what was reported, the model produced separate estimates for patient QALYs and caregiver QALYs.

4. The DSU evaluation states that 1,000 patients are simulated. In fact, with the exception of the probabilistic sensitivity analyses, the results are based on 20,000 simulated patients in each treatment arm.

Data Used to Inform Decision to use a 20 week Cut-Point for Treatment Effect Estimates of Donepezil

Annualized Rate of Change in MMSE from Previous Visit
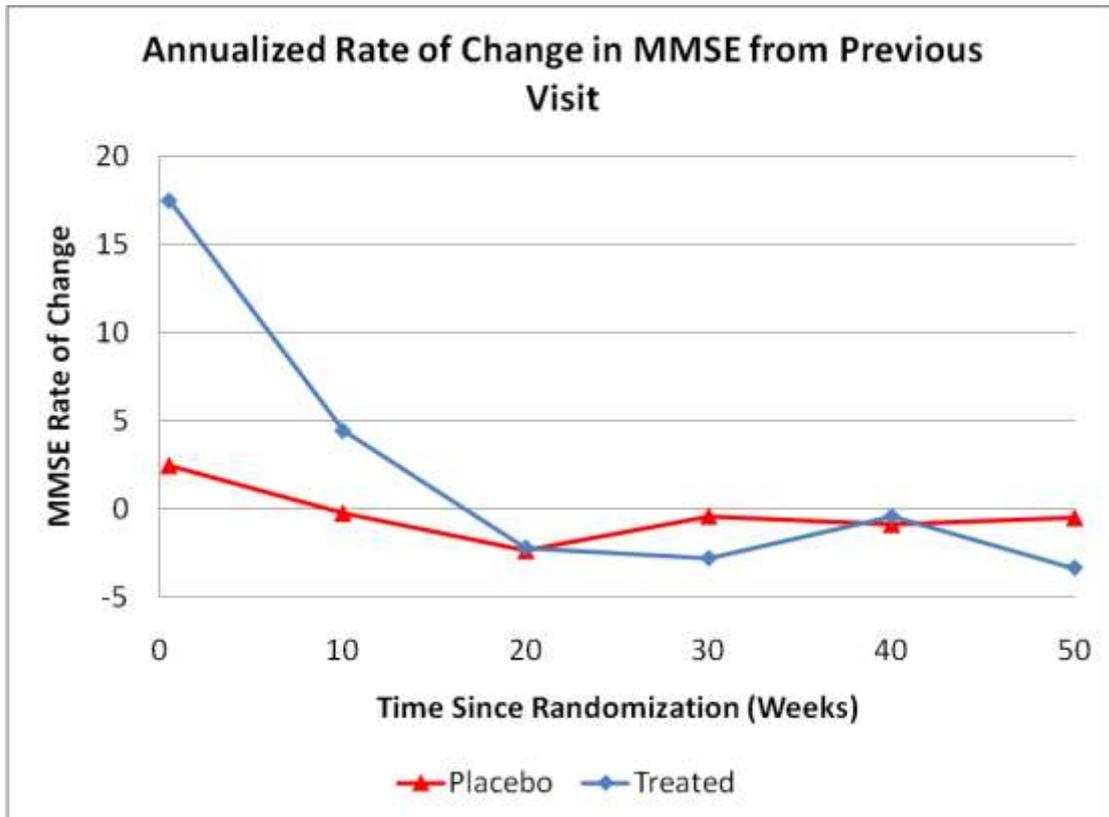
Appendix B: Additional Critique of the PenTAG Model

Lack of Face Validity on Predicted Outcomes for Patients in the Pre-Institutional Care State

- The model developed by PenTAG purports to capture health utility and cost differences resulting from treatment while patients are pre-institutionalised. To test this, we plotted the average cost and average utility for surviving patients assigned to best supportive care and donepezil over time for the proportion of the cohort remaining in pre-institutional care. We did this separately for the three age subgroups modelled. The analyses below reveal that the model is almost completely insensitive to changes in disease severity in the pre-institutionalisation period.

- The first series of graphs is for health utilities. The graph plots the average health utility assigned to surviving patients in pre-institutional care over the model time horizon. No discernible difference in mean health utility can be seen on these charts although evaluation of the data that goes into the chart shows small differences. The mean utilities by age group are presented in the table below.

Age Group	Best Supportive Care (denoted as BST in graphs)	Donepezil (denoted as Don in graphs)
69	0.5777	0.5767
77	0.5406	0.5394
86	0.5009	0.4999

- Of greater concern is the lack of sensitivity of the model to changes in health utilities regardless of treatment. As the charts indicate, mean utilities for surviving patients do not change over 15 years in the model. If one accepts these results, one must also accept that over 15 years for patients who continue to reside in the community, disease progression has no influence on patient quality of life. The first change in mean health utility calculated in the model happens at year 16.2 in all age groups and for all treatments. We note that this change occurs after the error in the calculations identified where the PenTAG model starts referencing empty cells to calculate utilities and costs (year 14.1). If this error is corrected, utilities remain constant for the entire modelling time horizon.
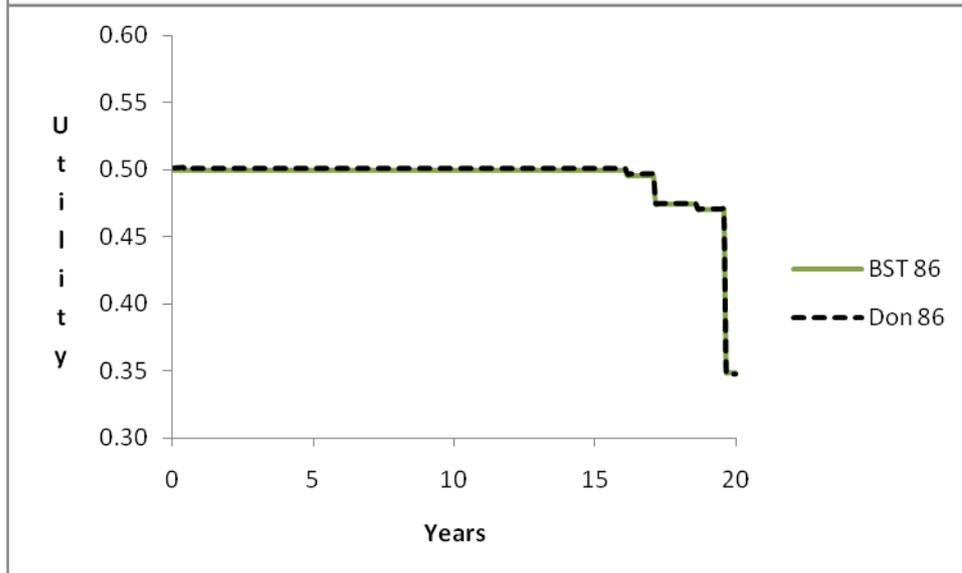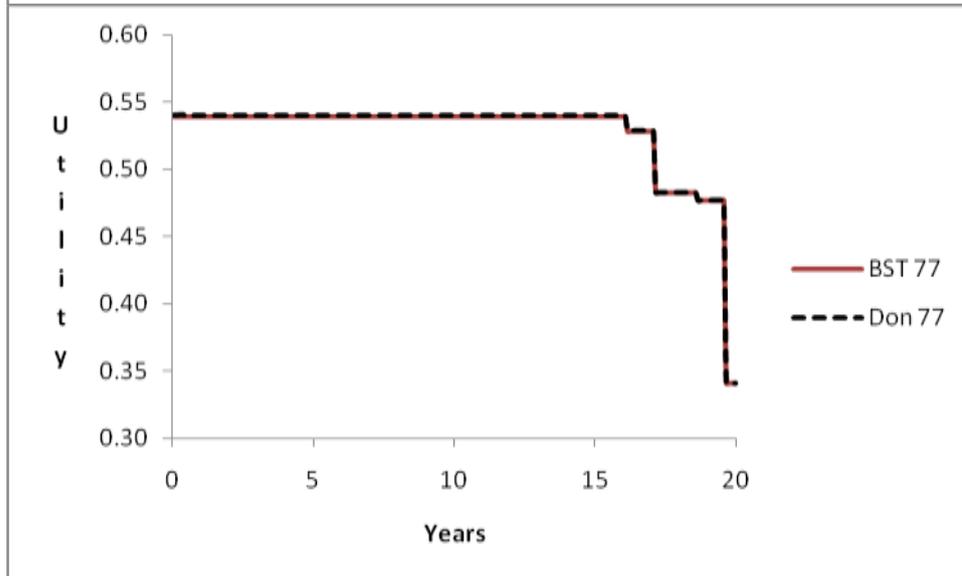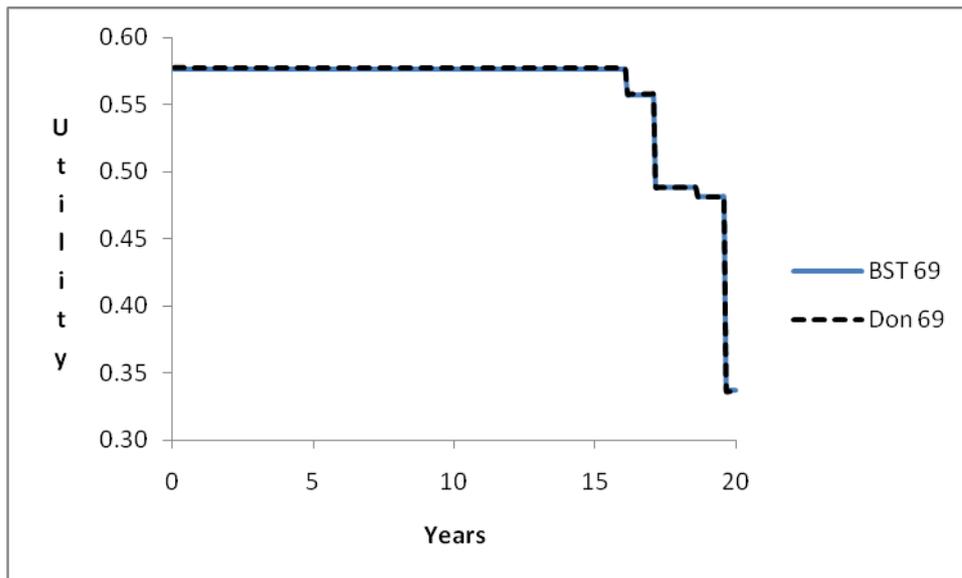
- As with utilities, the model also appears to be insensitive to changes in costs. The charts below plot the average cost assigned to surviving patients who are not in institutional

care over the modelling time horizon. The only differences observed in the charts below occur after 16 years (again, after the period where the model is referencing empty cells as part of the cost calculation). Again, there are actually some small differences between donepezil and best supportive care in the model as indicated in the table below which provides the average monthly cost of care in the community by age group and treatment.

Age Group	Best Supportive Care	Donepezil
69	£1,338	£1,332
77	£1,567	£1,559
86	£1,890	£1,881

- As with health utilities, the PenTAG model indicates that costs of care in the community for patients remaining in the community do not change for almost 15 years and not significantly until well after 15 years. The first change in costs in the PenTAG model occurs at year 14.2, which once again is after the PenTAG model starts referencing empty cells in its cost calculations. In fact, at year 14.2, because of the error, costs drop slightly and continue to fall until year 15.7. If this error is corrected, costs remain constant for the entire modelling time horizon.
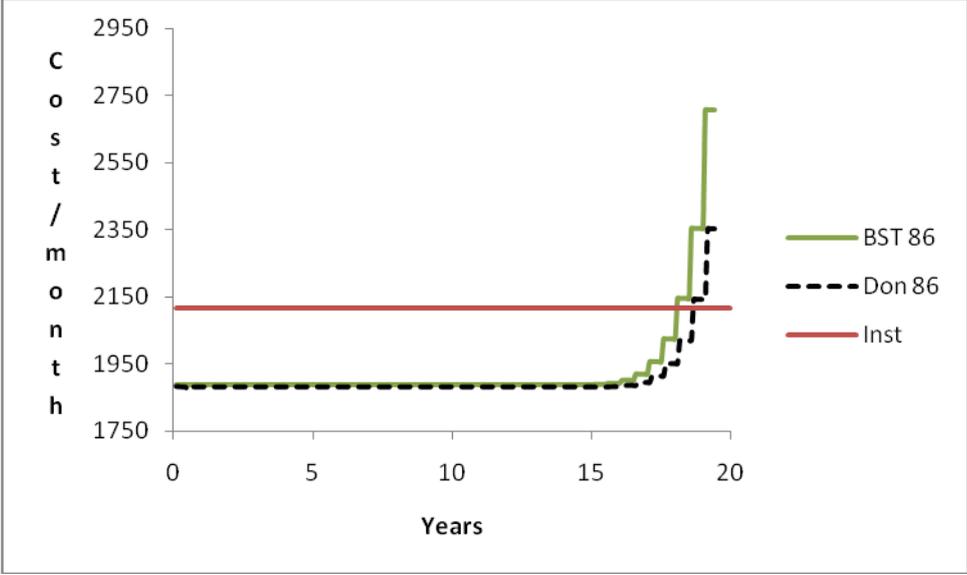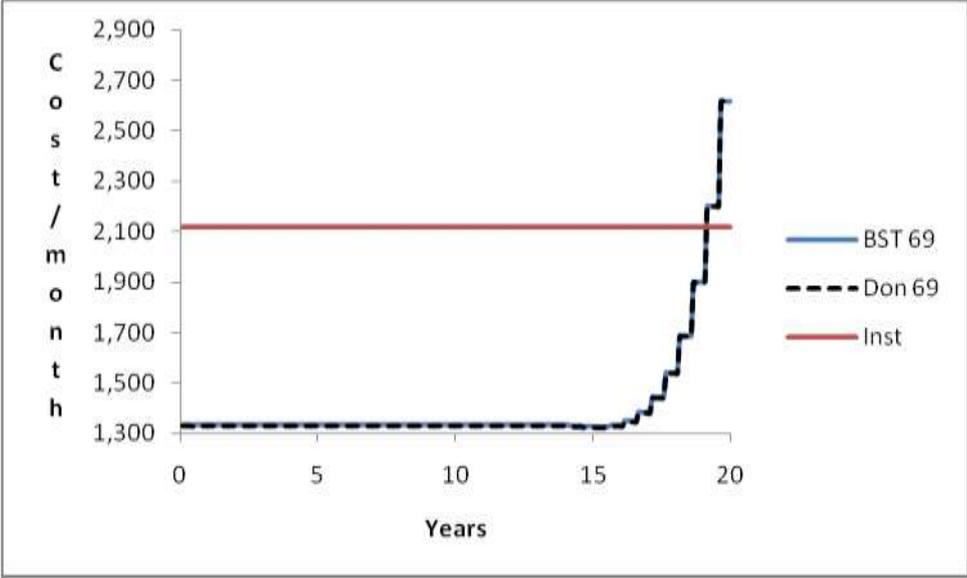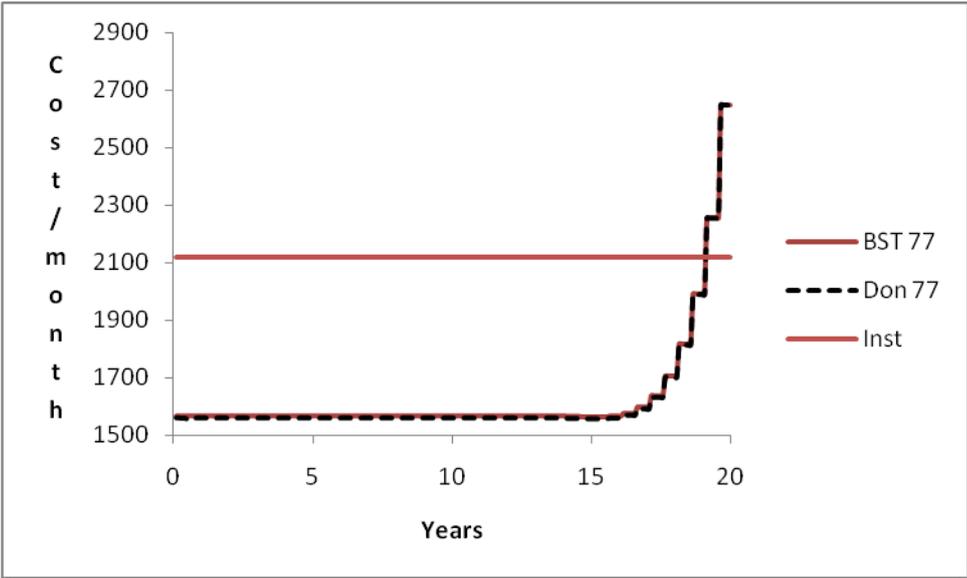
Similar plots can be produced for the implied average MMSE in the pre-institutionalized population, and indicate that mean MMSE in this cohort does not change for 15 years (again, MMSE scores only start to fall in the model because of a technical error). This occurs because the structure of the model implicitly assumes that the only determinant of disease severity while in pre-institutional care is time to institutionalization. Decreasing the risk of institutionalization would results in even longer periods over which MMSE does not change, whereas increasing the risk would shorten this time.

No Reduction in Risk of Institutionalization over Treatment Period

- Although some nominal benefits are allowed over the treatment period, the first 6 months in the model do not allow for any benefit in terms of delaying the time to institutionalisation. Incorporating longer term trial data would mean modelling a longer trial period, and thus a longer period where reducing the risk of institutionalisation is not modelled.

Proportion of Patients in Institutional Care at Start of Model is not Handled Appropriately

- Since the model assumes that treatment stops when patients enter institutional care, there seems to be no logic in modelling a population that starts in this setting. The rationale for doing so is that treatment in institutional care is not outside the license of use for cholinesterase inhibitors, which is a valid point. The inclusion of a subset of patients who start in institutional care, but by definition cannot be treated due to model assumptions, is not a valid response. Under the current model, the impact of changing the percent of patients who start the model in institutional care on ICERs is nil.

Model Does not Address Question of Treating Newly Diagnosed Cases of Alzheimer's Disease

- As, by admission from PenTAG, the model they developed is only valid for prevalent cases of Alzheimer's disease, and not new cases, the scope of the question addressed by PenTAG is indeed very limited, as it excludes all new cases of Alzheimer's disease in the UK and does not answer the question of whether these patients should be treated or not. We would also, however, argue that given the concerns expressed in this response the model has no validity for prevalent cases.

Assumption that All Patients in Institutional Care have MMSE Scores Below 10 is Incorrect

- The PenTAG model assumes that all individuals in institutional care have MMSE scores below 10, but the data used by PenTAG and cited in their report, indicate that a significant proportion of patients with MMSE above 10 are institutionalised, as do the LASER-AD data presented in the PenTAG report.

Treatment Discontinuation is not Correctly Modelled

- The PenTAG model purports to factor in treatment discontinuation. Unfortunately, it does so by modifying treatment costs only. While this favours outcomes with cholinesterase inhibitors, it leads to a model which predicts that the most favourable economic outcomes with treatment are attained when all patients stop treatment immediately.

Drug Costs are Incorrect or Inappropriate

- The cost used by PenTAG for donepezil 10mg of £3.18 is incorrect. The correct cost is £3.00 per day.
- Considering the proximity of the loss of patent exclusivity for donepezil in 2012, it would seem appropriate for PenTAG to have run scenario analyses based on projected changes in drug prices after the introduction of generic cholinesterase inhibitors. However, this was not done.
- The use of an average cost for galantamine and rivastigmine based on the distribution of doses used in the clinical trials is completely inappropriate. If a weighted cost had to be used, the appropriate weightings would be based on the doses used in actual practice. Consequently, treatment effects should also have been adjusted to look at actual doses used, or different dose levels should have been analysed separately.

Incorrect Data are Extracted from the Source Paper on Patient Health Utilities

- The description of the PenTAG model states that EQ-5D time trade off scores from Jonsson 2006 were used in the base case analyses for informing patient health utility inputs, but both the values in the report and those in the electronic model do not match the EQ-5D results from Jonsson 2006. Closer inspection of the source paper reveals that the data used by PenTAG are actually the average proxy-rated health utilities from several different measures, the EQ-5D time trade off scores, the EQ-5D visual analog scale (VAS) and the Quality of Life – Alzheimer's Disease (QOL-AD) instrument.

The SHTAC Model Results Should Not be Considered as Supporting the PenTAG Model Results

- Although the comparison of the PenTAG and SHTAC models starts by concluding that the cost-utility estimates are similar, we feel obliged to highlight vast differences in estimates produced by these two models:
  - Total costs of care for patients over their lifetime are 50% lower in the PenTAG model compared to the SHTAC model.
  - Total QALYs per patient are roughly 35% lower in the PenTAG model than the SHTAC model. QALYs gained with treatment compared to no treatment are over 80% lower.
  - Total time in institutional care is 64% lower in the PenTAG model.
  - The delay to institutional care associated with treatment is also over 80% lower in the PenTAG model.
- The two models in fact produce incoherent results, and the similarity in ICERs cannot be considered as supporting evidence for either model

Appendix C: Concerns with Statistical Analyses Conducted by PenTAG

A number of analyses have been conducted on the Oxfordshire data, and although reporting on the methods is extremely limited, there appear to be some serious problems with the analyses used to derive the equations that drive the model results.

We note that the equations developed by PenTAG directly contradict the findings of previous analyses based on this work (Wolstenholme 2002) but PenTAG has made no effort to explain why their results diverge so significantly from previous findings using the same data source.

Time to End of Pre-Institutionalisation

This is the single-most important equation in the PenTAG model, yet PenTAG opted to select an exponential distribution for the fit 'for simplicity'. PenTAG does not report how well the equation they derived fit the observed data, but from Figure 64, it appears that the models developed do not fit the observed data.

- Panel A in Figure 64, which shows observed versus predicted survival by high, medium and low MMSE scores at baseline shows that the fit for high MMSE scores, consistently and significantly overestimates the risk of exiting pre-institutionalisation. The fit for low MMSE scores, consistently and significantly underestimates this risk. In fact, the fit for low MMSE scores seems to match the observed data for 'medium' baseline MMSE scores much better than for low baseline MMSE scores.

- Panel B in Figure 64, which shows observed versus predicted survival by high, medium and low ADL scores at baseline shows that the fit for high ADL scores does not match the observed data at all, overestimating the risk of exiting pre-institutional care.

- Panel C in Figure 64, observed versus predicted survival by high, medium and low age at baseline also indicates a poor fit of the data, with the predicted outcomes not matching the observed data at all. In this case, it is clear that the functional form selected by PenTAG, does not even closely match the shape of the observed data.

Survival

Figure 65 indicates that the models used by PenTAG do not even remotely fit the observed data. The only similarity between the observed and predicted survival curves is that they are both downward sloping, which by definition they must be. Since according to the structure of PenTAG's model, these curves determine how long patients spend in institutional care, the poor fits completely compromise these calculations.

MMSE as a Function of Time to End of Pre-Institutionalisation

Figure 66 shows that the linear fits produced by PenTAG bear no resemblance to the observed data. Further the need to develop separate equations for mild to moderate and moderate to severe patients (defined by MMSE) is a clear indication that the models fail to accurately predict MMSE as a function of time to institutionalisation. If both models were accurate, they would have produced similar equations. As noted previously, the models predict MMSE scores well above 30, and do not allow scores below 5.2 to 8.3 (depending on which equation is used).