A REVIEW OF THE EISAI / PFIZER ECONOMIC MODEL ON THE COST-EFFECTIVENESS OF DONEPEZIL

REPORT BY THE DECISION SUPPORT UNIT

Praveen Thokala, Sarah Davis, and Allan Wailoo
School of Health and Related Research, University of Sheffield

8 June 2010

CONTENTS

1. INTRODUCTION 4
1.1. BACKGROUND 4
1.2. METHODS 4
1.3. STRUCTURE OF THIS REPORT 5
2. MODEL LOGIC 6
2.1. PATIENTS 6
2.2. UPDATES 7
2.3. EVENTS 10
3. CONCERNS AND ERRORS IN THE MODEL 11
3.1. CONCERNS 11
3.1.1. Patient population 11
3.1.2. Structure 12
3.1.3. Treatment benefits 13
3.1.1. Time to discontinuation of treatment 13
3.1.2. Error suppression in calculations 14
3.1.3. Redundancy 14
3.2. ERRORS IN THE MODEL 15
3.2.1. Life-expectancy 15
3.2.2. MMSE Scaling 16
3.2.3. Application of hazard ratios for discontinuation 16
3.2.4. Discrepancies between the model and the submission 18
3.3. PROBABILISTIC SENSITIVITY ANALYSIS 18
3.3.1. Beta distributions for institutional care 19
4. SUMMARY OF AMENDMENTS TO THE BASECASE MODEL 21
4.1. AMENDMENTS MADE TO THE BASECASE 21
4.1.1. MMSE Scaling 21
4.1.2. Life-expectancy 21
4.1.3. Hazard calculations 22
4.1.4. New Base Case 22
4.1.5. New PSA results 23
4.2. EXPLORATORY ANALYSES ON THE NEW BASECASE 24
4.2.1. Proportion Institutionalised 24
4.2.2. Impact of institutionalisation on caregiver utility 25
4.2.1. Potential overestimation of treatment effect 26
4.2.2. Combined effect of the exploratory studies so far 26
4.2.3. Regular update interval 27
4.2.4. Distribution of sampled life-expectancy estimates 28
5. CONCLUSIONS 30
6. REFERENCES 31

ABBREVIATIONS AND DEFINITIONS

ADL	Most recent ADL
AgeOrig	Age at Baseline
Age_cg	Age of the caregiver
Age	Current age of the patient
BaseADL	Base ADL
BaseIADL	Base IADL
BaseIADLWeeks	Base IADL x Weeks
BaseMMSE	Base MMSE
BaseNPI	Base NPI
BaseNPIWeeks	Base NPI x Weeks
CGU	Caregiver utility
DSU	Decision support unit
IADL	Most recent IADL
MMSE	Most recent MMSE
norm(a,b)	Normal distribution with mean a and standard deviation b
NPI	Most recent NPI
PM1	min(PrevMMSE, 9)
PM2	max[0, min(PrevMMSE-9, 9)]
PM3	max[0, min(PrevMMSE-18, 12)]
PrevMMSE	Previous MMSE
PrevNPI	Last NPI
PrevADL	Previous ADL
PrevIADL	Previous IADL
PrevMMSEChange	Previous Rate of MMSE Change
PsyMed	Use of Psychiatric Medications
PU	Patient utility
Tnow	Current time
Tup	Time of last update
Weeks	Time in weeks from start

1. INTRODUCTION

1.1. BACKGROUND

NICE is reviewing its existing technology appraisal guidance [TA 111] on the use of donepezil, galantamine, rivastigmine and memantine in the treatment of Alzheimer's Disease. The economic model included within Eisai/Pfizer's submission to NICE compares donepezil to no treatment using a discrete event approach and has been built using the modelling package ARENA. The Decision Support Unit was asked by the Assessment Group (PenTAG) to examine the executable model submitted by Eisai/Pfizer with the following aims;

1) To establish how the modelling methods described in the manufacturer's submission are implemented within the ARENA modelling software.
2) To detect any errors in the implementation of the model and to examine the potential impact of those errors on the results.
3) To conduct additional sensitivity analyses to explore any areas of the model where the approach used in the submitted model is not considered to be robust by the assessment group.

1.2. METHODS

The DSU examined the ARENA model by considering the entities, attributes, variables and expressions defined within the model, and the SIMAN logic which describes the flow of patients through the model and the processes that occur at each stage. The model pulls in data inputs from an Excel spreadsheet and this was examined to establish whether the data inputs match those reported in the submission. We made selected changes to the Excel data input file and verified that these resulted in appropriate changes in the model results. The model has several sub-models and these were examined to determine the function of each sub-model and to identify any redundant features that are not employed within the submitted economic evaluation. The general model logic was examined to determine the flow of patients from one sub-model to the next. The SIMAN logic was then examined in detail to establish how each aspect of the model described in the submission is implemented within the ARENA model.

1.3. STRUCTURE OF THIS REPORT

The second chapter of this report contains a general description of the model which is provided to supplement the details supplied within the manufacturer's submission. The third chapter of this report documents the factual errors which we have found within the ARENA model and the steps we have taken to correct these errors. The fourth chapter describes the results for the corrected basecase and several exploratory analyses that we have conducted to establish what the cost-effectiveness results would be if various changes were made to the model assumptions.

2. MODEL LOGIC

The Arena model submitted by the manufacturer is a generic model which has a variety of other modules/logic which are not relevant for the study undertaken. For example, it includes a screening module, and an option for patients to restart treatment as well as having the provision to estimate costs/utilities of two additional drugs along with donepezil and no treatment. Notwithstanding the model's capability to perform different analyses, this report will only comment on the issues in the model that are related to the cost-effectiveness analysis of donepezil against no treatment.

2.1. PATIENTS

The model utilises a weighted sampling approach to sample the patients in the model from the trial population. The trial population consists of 826 trial patients and there is a provision to select the patients based on different characteristics such as age, sex, MMSE, etc. The two main subgroups utilised are a mild patient group (221 patients with 20<=MMSE<=26) and a moderate patient group (542 patients with 10<=MMSE<20). The model utilises 1000 patients and these are sampled from the corresponding subgroup utilising a weighted approach i.e. if using a mild population, 1000 patients are sampled from the 221 mild patients and are assigned the characteristics of the corresponding trial patient. These characteristics include age, sex, race, MMSE, NPI, ADL, IADL, previous MMSE and the change in MMSE in the previous year, as well as other information such as whether they are on psychiatric medications, whether they are living with their primary caregiver and if so, the caregiver's age and gender. These characteristics are specific to the individual patients and are assigned to patients as attributes.

As there are fewer patients in the trial population, than in the sampled model population, it is likely that the same trial patient with be included more than once in the modelled population. As the sampling is weighted to achieve an age and sex distribution that is consistent with the UK AD population, this may mean that some patients whose characteristics are rare in the trial data set, but common in the UK AD population, may be sampled multiple times and their individual characteristics may have a disproportionate influence on the overall results.

The patients are then cloned i.e. each patient is separated into two identical patients with the exact same characteristics. One of the hypothetical patients is then allocated to the donepezil arm of the model and the other is allocated to the no treatment arm.

2.2. UPDATES

The attributes of each patient are updated at different time intervals in order to replicate the progression of the disease and are then used by the model to perform cost effectiveness analysis. The model keeps track of four disease measures; MMSE, NPI, ADL and IADL. It should be noted that the MMSE equation uses annual increments while the other three equations use time as a continuous variable to estimate the new values.

The underlying progression equation for MMSE is defined as follows;

MMSE = MMSE+ (- 5.4663 + norm(0,0.5) - 0.4299PM1 - 0.0042PM2 + 0.1415PM3 - 0.0791PrevMMSEChange + 0.0747Ageorig)(Tnow-Tup)/365.25

MMSE_trtmnt = MMSE + (T_eff - 5.4663 + norm(0,0.5) - 0.429PM1 - 0.004PM2 + 0.1415PM3 -0.079PrevMMSEChange + 0.0747Ageorig)(Tnow-Tup)/365.25

where T_eff is 6.1583 if time is less than 20 weeks and 2.4671 otherwise. The treatment effects only last for one year after which it is assumed to be zero.

The underlying NPI equation in the model is defined as

NPI=(BaseNPI+5.74+norm(0,3.75) +0.03Weeks-0.59BaseNPI-0.0012BaseNPI*Weeks +0.24PrevNPI-1.74White-3.82Black+2.34PsyMed+0.12BaseMMSE-0.22MMSE- 0.64if treatment)*1.44

The underlying ADL equation in the model is defined as

ADL=BaseADL+1.35+norm(0,2.48)+0.06Weeks-0.79BaseADL+0.71PrevADL+ 0.12BaseMMSE+0.09Ageorig+0.81PsyMed-3.05White-0.49MMSE-0.81iftreatment

The underlying IADL equation in the model is defined as

IADL=BaseIADL+1.27+norm(0,1.9)+0.17Weeks-0.84BaseIADL-0.002BaseIADL*Weeks+0.84PrevADL-0.67Male+0.20BaseMMSE-0.28MMSE-0.16BaseADL+0.18ADL+if treatment, then add (0.63-0.062*Weeks)

The term norm(0,x) appearing in each of the disease progression equations is a random intercept parameter which is included to introduce patient level variation to the disease progression. This random variation is in addition to the variation provided by each patient having unique characteristics.

The patients are assigned a severity level based on their MMSE scores after every update. The severity categories and their MMSE ranges are shown in Table 1. The time spent in different severity levels are accumulated for all the patients in the donepezil arm as well as the no treatment arm. The proportion of patients in institutional care is dependent on the severity level as shown in Table 1.

Table 1: Proportion of patients in institutional care according to severity level

MMSE	**Severity scale**	**Home**	**Institutional Care**
25 to 30	Mild	87.1%	12.9%
20 to 24	Mild Moderate	74.4%	25.6%
15 to 19	Moderate	61.7%	38.3%
10 to 14	Moderate Severe	49.0%	51.0%
0 to 9	Severe	30.0%	70.0%

Even though the model utilises an individual patient approach, the patient and caregiver utilities are estimated using average values. For example, in the patient utility equation the "Institutionalised" covariate is, strictly speaking, a factor or dummy variable that takes a value of 1 if the patient is institutionalised and 0 if not. The cost effectiveness model does not classify individual patients as institutionalised or not. Rather they are assigned a probability of being institutionalised based on their MMSE score.

PU = 0.408 + 0.010MMSE - 0.004NPI - 0.159Institutionalised + 0.051Living with carer

CGU = 0.90 - 0.003Age_cg + 0.03Male + 0.001Age - 0.001NPI - 0.001ADL -0.0004IADL + 0.01PsyMed

The QALYs for both donepezil treated and untreated patients are accumulated by multiplying the utility of each individual patient by the time they spent in that state. The model estimates both discounted and undiscounted values of QALYs.

The costs for both donepezil treated and untreated patients are estimated by accumulating the treatment costs (for patients under treatment) and the patient care costs for home or institutional care. These monthly patient care costs are based on severity level as seen in Table 2. Again, although the model is based on an individual patient approach, patient care costs are estimated by multiplying the weighted averaging based on severity level by the time spent in that severity state. Drug treatment costs are accrued according to the number of days on treatment. In addition to the drug treatment costs, patients also incur the cost of a medical consultation every 6 months whilst on treatment.

Table 2: Monthly patient costs according to severity level and location of care

Severity	Monthly Medical Costs £(Home)	Monthly Medical Costs £ (Institutional)
Mild	687	2801
Mild Moderate	742	2801
Moderate	798	2801
Moderate Severe	878	2801
Severe	957	2801

PatientCareCosts = (Probability of home care*monthly home medical costs + Probability of Institutional care*Monthly Institutional Costs)*(Tnow-Tup)*12/365.25

DrugTreatmentCosts = TmtCosts*(Tnow-Tup)

The caregiver times are estimated by the model but the caregiver costs are not taken into account. Hence, the total costs are calculated by adding the treatment costs to the patient care costs and the model estimates both discounted and undiscounted values of total costs. The discounted and undiscounted costs accumulated in different severity levels are also calculated.

2.3. EVENTS

Patient characteristics are updated and the costs along with QALYs are calculated every time the patient undergoes an event. The events that occur in the life of a patient and the times when they occur are presented in Table 3.

Table 3: Events occurring in the life of a patient

Event	Time (in days)
Start Treatment	0.01
Checks for Discontinuation*	0.02, 91.3, 182.6 and 365.25
Regular updates	Every 3 months
MD Visit	Every 6 months while in treatment
Stop Treatment	Patient specific time
Death	Patient specific
Last update/Model End	9131.25

*this event is used to assign a new Stop Treatment time

The patients undergo all but two events at exactly the same time; it is only the time to stop treatment and time to death that are individual to each patient. The submodel "sMain" in the model searches for the next event for every patient and sends them to the earliest event. Each event is a submodel in which the patients are processed after which they are returned to the submodel "sUpdater". The updater submodel calculates the new disease measures and estimates the costs and QALYs.

3. CONCERNS AND ERRORS IN THE MODEL

This section is intended to summaries the issues we identified while examining the model. The first part of this section contains a critique of the model which is provided to highlight the key issues within the manufacturer's submission. The second part of this section documents the factual errors which we have found within the ARENA model. It should be noted that due to the large amount of redundant code in the model, it was very difficult to perform an independent review of the model. Despite the lack of transparency in the manufacturer's model, the DSU was able to identify the key parts of the model. This chapter describes the concerns and errors within these key areas.

3.1. CONCERNS

3.1.1. Patient population

The modelled population is sampled from trial patients but it is weighting by age and sex to match the distribution of these variables in the UK AD population. The weighted sampling is done from the patient population after it has been filtered to include only mild or only moderate patients. It should therefore produce age and sex distributions that are similar in each severity category. However, the simulated moderate population has a better mean survival than the simulated mild population (4.603 vs. 4.110 years). The manufacturer states, in section 3.4.14 of their submission, that this is because the simulated moderate population is younger and has a higher proportion of females. This may produce misleading results if

patients with mild disease are actually more likely to be younger than patients with moderate disease in the UK AD population. It also suggests that the method used to weight the sampling to match the age and sex distribution in the UK is not functioning effectively. This could be because there are insufficient patients in the data set from which the population is sampled as previously discussed in section 2.1.

The DSU is also concerned that other characteristics of the sampled population may not match the UK AD population e.g. likelihood of living with carer, use of psychiatric meds, ethnicity, etc. For example, everyone lives with the carer in the sampled patient population, which is unlikely to be representative of UK patients. We have been unable to investigate the sensitivity of the model to changes in the patient characteristics due to the way the model samples its patient population from the trial data,

3.1.2. Structure

The DES approach has been used to track multiple patient characteristics, but these are updated at fixed intervals (e.g. 3mths). In a Markov model, a half-cycle correction would be applied to estimate the costs and QALYs based on the distribution of patients across the health states at the midpoint of each time-cycle. In this DES model, we effectively have a 3 month time-cycle but no equivalent "half-cycle type" correction is applied. Therefore if the time since the last update is 3mths, then the costs and utilities applied during those three months are based on patient variables at the end of the three months.

Even though it is claimed that this is a DES approach, the model calculates two of the most important parameters in determining costs and effects (patient care costs and utilities) using weighted averages in the same manner as a cohort model. Location of care (Home or Institutional) is not modelled on an individual level but is based on the mean rate for patients according to severity. The model is not a pure DES type model but incorporates elements of individual sampling and cohort modelling approaches.

3.1.3. Treatment benefits

The MMSE treatment effect is modelled as an "absolute" benefit rather than using relative risk methodology. i.e. for each period on treatment there is a fixed, absolute change in MMSE. This assumes that all patients receive the same benefit (mean value derived from trial population) irrespective of their characteristics such as severity, age, sex, etc. This absolute benefit is then applied to the untreated progression which is based on the CERAD data.

It is often assumed when building models that the relative risks from a trial are independent of the baseline risks and can therefore be applied to baseline risks estimated from cohorts which may be more representative of the population being treated. However, we are concerned about the approach taken in this situation. It is questionable that the reduction in progression achieved by treatment and estimated from the trial data is independent of the underlying rate of progression. MMSE treatment effect is one of the key drivers in this model, so if the absolute treatment effect is not transferable from the trial patients to the CERAD cohort patients, this could have substantial effects on the estimated ICERs.

Furthermore, there is also the potential for double counting of the treatment effect. Whilst treatment affects both MMSE and NPI directly as covariates in the regression equations, there is also an additional link between the two measures since MMSE is also a covariate in the NPI regression.

3.1.1. Time to discontinuation of treatment

Different discontinuation rates are applied for different time periods within the model. The rates are presented in Table 8 of the manufacturer's submission as fixed probabilities over discrete time periods. In the model, it is assumed that the hazard is constant over each of these discrete time periods, allowing the hazard to be calculated from an exponential survival distribution. The hazard is then adjusted for three continuous risk factors which increase the risk of discontinuation. The individual's time to discontinuation, T_d, is then sampled using;

$$T_d = - LN(UNIF(0,1)) / \text{adjusted hazard.}$$

This time to discontinuation is re-sampled at the start of each discrete time period (0, 3, 6, 12mths). Each time a new sample is taken from the uniform distribution meaning that an individual who is sampled to have a higher than average risk of discontinuation in the first time interval (0 to 3 months) can then be sampled to have a lower than average risk of discontinuation in the next interval (3 to 6 months) even before the discontinuation risk has been adjusted to account for their individual risk factors. Using the same sample from the uniform distribution for each time interval would allow the risk of discontinuation to be estimated more consistently for the individual over the course of their lifetime, but still allow the hazard to be updated according to changes in their risk factor profile during the first year.

3.1.2. Error suppression in calculations

There is an extensive use of various functions such as MIN, MAX, etc to suppress any implausible values that arise during calculations. For example, in utility and MMSE calculations (MX(0, utility)), MN(30,MMSE) and other similar expression are used to suppress any negative utilities values or any MMSE values greater than 30. It would have been better for any implausible values predicted by the model to have been recorded as errors and investigated rather than being suppressed in the calculations.

3.1.3. Redundancy

As mentioned earlier, the model submitted by the manufacturer is a generic model which has a variety of other modules/logic which are not relevant for the study undertaken. This redundancy is present throughout the model, which makes the review process very difficult and time-consuming. For example, although the utility equation in the model is correctly implemented, it is defined as a combination of five different equations. This lack of transparency means it is almost impossible to be certain that all the issues in the model have been identified.

3.2. ERRORS IN THE MODEL

3.2.1. Life-expectancy

The manufacturer's submission states that expected survival was calculated by fitting functions of the form;

Survival (years) = A x (Percent Surviving) ^ B

to the MRC CFAS data. The median survival estimates from the MRC CFAS data are given in Table 10 of the manufacturer's submission and the A and B parameters for males and females according to their age group are given in Table 11 of Appendix H to their submission.

The model samples the life-expectancy of the patients as follows;

Time to death (in years) = A x UNIF(0,1) ^ B

where A and B are taken selected from Table 11 of Appendix H for the appropriate age and gender of the patient. The DSU team identified the following mistakes in estimating the life expectancy of the patients in the model.

3.2.1.1.Male survival estimates are applied to females in one age category

For females with ages 70 to 79, the expression (eTimeEvDeath) which is being used to select the appropriate A and B is referring to the data for males rather than females. The model is therefore underestimating survival in this group as median survival is greater for females in this age category. This error affects both treated and untreated patients.

3.2.1.2.No survival estimate for age 90.

The expression (eTimeEvDeath) which is used in the model to select the appropriate A and B values according to age and gender defines the oldest age category as age>90 rather than age>=90. It therefore does not generate an expected survival for patients aged 90. This effectively set the expected survival to zero for patients who start the model with age =90. This error will essentially remove some patients from the model before they incur any costs

or accrue any QALYs and again, it affects both treated and untreated patients. There are four patients aged 90 in the set of 826 trial patients from which the modelled population is sampled, but it is unclear how many times these patients are included within the sampled population.

3.2.2. MMSE Scaling

The PrevMMSEChange term used in estimating the updated MMSE (equation) is the annual rate of change and therefore the change since the last update has to be scaled to give an annual rate. This is calculated in the model as;

PrevMMSEChange=(MMSE - PrevMMSE)/(365.25/(TNOW-aLastUpdate))

We believe that it should be calculated as;

PrevMMSEChange=(MMSE - PrevMMSE)*(365.25/(TNOW-aLastUpdate))

Given that updates usually occur at 3 monthly intervals, the PrevMMSEChange scores is being underestimated by factor of sixteen.

3.2.3. Application of hazard ratios for discontinuation

The hazard for discontinuation of treatment is adjusted for three risk factors which increase the risk of discontinuation. These risk factors are baseline MMSE, current MMSE, and annualized change in MMSE. The hazard ratios for these risk factors are specified for different time periods during the first year. The hazard ratios for these risk factors are only applied during the first year of treatment and are then set to unity. In Appendix H, it is stated that a Cox regression model was used to estimate the hazard ratios. In a Cox regression model, the natural logarithm of the hazard is assumed to be a linear function of the form;

$$LN(hazard) = beta_0 + beta_1 x_1 + beta_2 x_2 + beta_3 x_3$$

Given that the risk factors included within the analysis are all continuous variables, it would be usual to present either the regression coefficient (beta) or the hazard ratios (HR) for an

increase in 1 unit along the scale of the continuous variable (HR = exp(beta)). It can be seen from the equation above that the hazard ratio for a decrease in one unit is the reciprocal of the hazard ratio for an increase in one unit. Likewise, the hazard ratio for an increase in two units is the square of the hazard ratio for an increase in one unit. More generally, if HR_1 is the hazard ratio for an increase in 1 unit from the reference range, then the hazard ratio for y units difference from the reference range is defined as follows;

$$HR_y = HR_1 \wedge y$$

In the model, the expressions aHRb, aHRc and aHRr are used to calculate the hazard ratio for the patient's baseline MMSE, current MMSE and annualised change in MMSE, as compared to the reference range for each of these variables. However, these are not being calculated in a manner which is consistent with a Cox proportional hazards model. Instead the following is being calculated;

HR_1 = hazard ratio for 1 unit increase in MMSE

For y > reference range: $HR_y = (HR_1-1)*y+1$

For y < reference range: $HR_y= (HR-1)*(1/y)+1$

Table 4. Reference ranges used for the continuous risk factors.

Risk factor	Months 0 - 3	Months 3 -6	Months 6-12	After 12 mths
Baseline MMSE	18.8	18.8	18.8	1
Current MMSE	19.3	18.8	17.8	1
Annualised change in MMSE	4.31	-2.15	-2.69	1

Therefore, in the model MMSE scores which are lower and higher than the reference range both increase the risk of discontinuation rather than lower ones decreasing the risk and higher ones increasing the risk. The reference ranges used in the model are given in Table 4 for information as these are not reported in the manufacturer's submission.

3.2.4. Discrepancies between the model and the submission

We have identified five instances where the data in the manufacture's submission does not match that being used in the model. The differences found were as follows;

a) The constant in calculating the annual rate of decline in MMSE is -5.4663 in the model calculations instead of 5.4663 as mentioned on page 89

b) In the NPI equation, the coefficient for the interaction term, baseNPI x weeks, in the model is -0.0012 instead of -0.59 as reported in NICE submission (page 90 of the submission). The same coefficient is reported as 0.0012 in Appendix H, table 5.

c) The coefficient for the interaction term, baseIADL x weeks, in the IADL equation is -0.002 in the model instead of 0.002 as mentioned in table 7 of appendix H.

d) The caregiver utility equation uses 0.013 as coefficient for PsyMed instead of -0.01 in the report (pg 93 of the submission) and in Table 15 of Appendix H.

e) The caregiver utility equation has a patient age term with a coefficient of 0.0014 in the model. Also, it has no term for patient gender as reported in pg 93 of the submission)

The first four discrepancies listed above were confirmed by the manufacturer to be typographical errors in the report and therefore do not alter the reported results. The fifth affects utilities of treated and untreated patient equally and therefore does not affect the incremental cost-effectiveness.

3.3. PROBABILISTIC SENSITIVITY ANALYSIS

The DSU replicated the probabilistic sensitivity analysis as detailed in the report i.e. 350 runs with 5000 patients each run for mild and moderate patients separately. Jackknifing [1] has

been performed on the results to identify the confidence intervals and they are reported in Table 5.

Table 5. Jackknifing analysis on manufacturer's model PSA (350 runs)

	Deterministic			Stochastic		
	Cost £	**QALYs**	**ICER £**	**Cost £**	**QALYs**	**ICER £ (95% CI)**
Mild	-3,386	0.147	-22,975	-1,786	0.130	-13,764 (-18,873 to -8,768)
Moderate	-1,883	0.109	-17,310	-1,316	0.105	-12,585 (-17,727 to -7,553)

From the table, it can be observed that the deterministic mean is quite different to the stochastic mean even though all the ICER's indicate that donepezil dominates standard care. In fact, the mean cost savings and QALY gains are smaller in the PSA analysis for both mild and moderate patients which means that the base case results presented in the submission are quite optimistic relative to the probabilistic mean. This would suggest that the deterministic ICERs cannot be used as a good estimate of the expected cost-effectiveness and that a robust PSA analysis is needed.

However, we have concerns regarding the implementation of the PSA analysis. In the health utility equations, all of the terms in the equation are varied within the PSA but each term is allowed to vary independently of the others removing any correlation between the terms. For the disease progression equations, only the intercept term and the treatment effects are varied within the PSA analysis. Again this removes any correlation between the intercept term and the other terms which are fixed. We also have specific concerns regarding the beta distributions used to describe the probability of institutional care as described below. The results of the PSA analysis should therefore be interpreted with caution.

3.3.1. Beta distributions for institutional care

The model uses beta distributions to describe the uncertainty in the proportion of patients receiving institutional care for each severity state. The alpha and beta parameters used to define the beta distribution are <1 for all severity states and are similar, but not exactly equivalent, to the average proportions in home and institutional care used in the deterministic analysis. When the alpha and beta parameters are both <1, this produces a U shaped beta distribution with asymptotes at 0 and 1 which does not seem to be a realistic distribution for

this parameter. No details are provided on how the alpha and beta parameters, which are given in Table 6 below, have been derived.

Table 6. Beta distribution for institutional care used in the submitted model

MMSE	Living in the community	Institutionalised	Distribution used in manufacturer model
Mild	87.1%	12.9%	Beta(0.86229,0.12771)
Mild-Moderate	74.4%	25.6%	Beta(0.73656, 0.25344)
Moderate	61.7%	38.3%	Beta(0.61083,0.37917)
Moderate-Severe	49.0%	51.0%	Beta(0.4851,0.5049)
Severe	30.0%	70.0%	Beta(0.297,0.693)

The following procedure is given in appendix H for calculating standard errors where these are not available from the literature. "Where a standard error was not available, we used ±25% of the parameter mean to assign a 95% confidence interval and calculate the corresponding standard error estimate." Using this method, we have calculated the 95%CI for the proportion receiving institutional care and used these to derive alpha and beta parameters for the proportion receiving care at home as shown in Table 7.

Table 7. Beta distribution for institutional care used in the DSU analysis

MMSE	Living in the community	Institutionalised (95% CI)*	Distribution used in DSU analysis for proportion living in community
Mild	87.1%	12.9% (9.7% to 16.1%)	Beta(360.6,53.4)
Mild-Moderate	74.4%	25.6% (19.2% to 32.0%)	Beta(132.2,45.5)
Moderate	61.7%	38.3% (28.7% to 47.9%)	Beta(60.5,37.5)
Moderate-Severe	49.0%	51.0% (38.3% to 63.8%)	Beta(28.4,29.6)
Severe	30.0%	70.0% (52.5% to 87.5%)	Beta(7.6,17.7)

*calculated as proportion +/-25%

These revised beta functions have been used to generate PSA estimates for the new base in chapter 4. However, this does not correct the other problems identified with the PSA described above. The results of the PSA analysis should therefore be interpreted with caution.

4. SUMMARY OF AMENDMENTS TO THE BASECASE MODEL

The model submitted by the manufacturer is updated to correct the errors mentioned in section 3.2 to create a new base case but the individual effects of each error correction are also provided here for the sake of completeness. The instances where the data in the report did not match that being used in the model (see section 3.2.4 above) were sent to the manufacturer for clarification; the manufacturer confirmed that the model parameters are correct and hence, they are left unchanged.

4.1. AMENDMENTS MADE TO THE BASECASE

4.1.1. MMSE Scaling

The PrevMMSEChange term used in estimating the updated MMSE (equation) is estimated using the correct scaling factor. The costs and QALYs estimated using the updated equation are presented against the basecase model in Table 8 below based on a deterministic ICER after 20 runs with 1000 patients.

Table 8. Cost effectiveness of base case model with corrected MMSE scaling

	Base case model			**Base case with corrected MMSE scaling**		
	Cost £	**QALYs**	**ICER £**	**Cost £**	**QALYs**	**ICER £**
Mild	-3386	0.147	-22975	-2953	0.137	-21554
Moderate	-1883	0.109	-17310	-1612	0.102	-15813

4.1.2. Life-expectancy

We modified the expression eTimeEvDeath to include patients aged 90 in the fourth age category and to select the appropriate estimates for A and B for females aged 70 to 79. The

impact on results from these two combined changes is seen in Table 9 based on a deterministic ICER after 20 runs with 1000 patients. The increase in cost-effectiveness of donepezil can be attributed to using the correct life expectancy for women, which was underestimated in the base case.

Table 9. Cost effectiveness of base case model with corrected life expectancy

	Base case model			**Base case with correct life expectancy**		
	Cost £	**QALYs**	**ICER £**	**Cost £**	**QALYs**	**ICER £**
Mild	-3386	0.147	-22975	-4118	0.178	-23125
Moderate	-1883	0.109	-17310	-2022	0.117	-17296

4.1.3. Hazard calculations

The new costs and QALYs are calculated by amending the expressions aHRb, aHRc and aHRr to use in the correct method for calculating the hazard ratios. The effect on results from these changes is seen in Table 10 below:

Table 10. Cost effectiveness of base case model with corrected hazard calculations

	Base case model			**Base case with correct hazard calculations**		
	Cost £	**QALYs**	**ICER £**	**Cost £**	**QALYs**	**ICER £**
Mild	-3386	0.147	-22975	-3345	0.146	-22960
Moderate	-1883	0.109	-17310	-1922	0.110	-17417

4.1.4. New Base Case

The new base case model is obtained by correcting the three errors identified in the manufacturer's model simultaneously. The model is 20 runs with 1000 patients (for both mild and moderate categories) and the ICERs are presented in Table 11 below

Table 11. Cost effectiveness of the new base case model

	Base case model			**New base case**		
	Cost £	**QALYs**	**ICER £**	**Cost £**	**QALYs**	**ICER £**
Mild	-3386	0.147	-22975	-3563	0.164	-21713
Moderate	-1883	0.109	-17310	-1763	0.111	-15824

4.1.5. New PSA results

These results incorporate the revised beta functions described in section 3.3.1 in addition to the amendments made to the basecase to produce the deterministic results in 4.1.4. The model was run 350 times for 5000 patients and jackknifing was performed to calculate the confidence intervals. It can be observed from Table 12 that the confidence interval is smaller for the new base case and this can be attributed to the fact that it uses the updated beta functions. The deterministic value is still towards the lower end of the interval obtained through the PSA analysis.

Table 12. Deterministic and PSA results for the manufacturer's base case and the new base case with corrected Beta distributions (350 runs)

	Deterministic			**Stochastic**		
	Cost £	**QALYs**	**ICER £**	**Cost £**	**QALYs**	**ICER £ (95% CI)**
Mild						
Base case model	-3,386	0.147	-22,975	-1,786	0.130	-13,764 (-18,873 to -8,768)
New base case model	-3,563	0.164	-21,713	-3,166	0.156	-20,282 (-22,837 to -17,730)
Moderate						
Base case model	-1,883	0.109	-17,310	-1,316	0.105	-12,585 (-17,728 to -7,553)
New base case model	-1,763	0.111	-15,824	-1,380	0.109	-12,678 (-15,309 to -10,057)

The model was also run 1000 times with 5000 patients to gather more accurate results and these are presented in Table 13. It can be observed that the confidence interval is smaller when using 1000 PSA samples rather than 350 PSA samples. Also, the stochastic mean is closer to the deterministic mean. These results suggest that it is necessary to run more than 350 samples to obtain an unbiased estimate using the PSA analysis.

Table 13. Deterministic and PSA results for the new base case with revised Beta distributions (350 & 1000 runs)

	Cost £	QALYs	ICER £ (95% CI)
Mild			
Deterministic	-3,563	0.164	-21,725
PSA with 350 samples	-3,166	0.156	-20,282 (-22,837 to -17,730)
PSA with 1000 samples	-3,415	0.159	-21,433 (-22,354 to -20,515)
Moderate			
Deterministic	-1,763	0.111	-15,882
PSA with 350 samples	-1,380	0.109	-12,678 (-15,309 to -10,057)
PSA with 1000 samples	-1,703	0.111	-15,285 (-16,686 to -13,888)

4.2. EXPLORATORY ANALYSES ON THE NEW BASECASE

4.2.1. Proportion Institutionalised

The proportion of patients institutionalised is dependent on the severity level as shown in Table 1. The patients are assigned a severity level based on their MMSE scores alone. PenTAG expressed concerns regarding the strength of MMSE as a predictor of institutionalization, so we performed exploratory sensitivity analysis assuming that the severity level has no effect on the probability of institutionalisation. This is implemented in the model by having the same proportion of patients institutionalised (36.5%) at all severity levels. This value of 36.5% is reported by the manufacturer as the overall percentage of institutionalised patients in the UK. The costs and QALYs are calculated and presented in Table 14 below:

Table 14. New Basecase with fixed institutionalisation across severity levels

	New Basecase			New Basecase with proportion of institutionalised patients set to 36.5% across all severity categories		
	Cost £	**QALYs**	**ICER**	**Cost £**	**QALYs**	**ICER £**
Mild	-3563	0.164	-21713	2186	0.113	£19389
Moderate	-1763	0.111	-15824	1826	0.077	£23676

Assuming MMSE has no effect on institutionalisation, the ICERs for the mild and moderate populations have become £19339 per QALY and £23676 per QALY, respectively.

4.2.2. Impact of institutionalisation on caregiver utility

In the model, caregiver utility is calculated using an equation which includes caregiver age and gender, the four main patient disease measures (MMSE, NPI, ADL, IADL) and use of psychiatric medicine. It does not contain any terms that relate to whether the carer is living with the patient and providing care in the home or whether the patient is living in an institution. The manufacturer's submission states that the caregiver utility equation has been derived using data from the Nordic, 324 and 312 trials, which are the same trials used to provide the patient data set from which the modelled population is sampled. Looking at this data set we find that all of the patients have the variable "living with patient" set to 1 suggesting that all patients had a caregiver living with them at the start of the study. They also state that information was not available on the impact of institutionalisation on caregiver utility. Therefore, caregiver utility is estimated in the model to be the same regardless of whether the caregiver is living with the patient and regardless of whether the patient is receiving home care or institutional care.

Treatment reduces progression to more severe disease states which are associated with a higher risk of institutional care in the model. If institutional care is associated with an increase in carer utility due to a lower burden of care being placed on the primary caregiver, then reducing disease progression and lowering the average time spent in institutional care will reduce expected QALYs for the caregivers. We investigated the sensitivity of the model to alternative assumptions regarding caregiver utility by removing the utility decrement associated with NPI, ADL and IADL for patients receiving institutional care. This improved

caregiver utility in both arms of the model, but the incremental effect of treatment on caregiver utility became negative as treatment delays institutionalisation which is associated with gains in care giver utility. The incremental costs and QALYs are calculated and presented in Table 15 below.

Table 15. New Basecase with modified caregiver utility

	New Basecase			New Basecase with improved carer utility after institutionalisation			
	Cost £	Carer QALYs	Total QALYs	Cost £	Carer QALYs	Total QALYs	ICER £
Mild	-3563	0.016	0.164	-3563	-0.010	0.138	-25844
Moderate	-1763	0.011	0.111	-1763	-0.010	0.091	-19399

4.2.1. Potential overestimation of treatment effect

In the manufacturer's model the NPI, ADL and IADL expressions have an MMSE term as well as having a treatment benefit term. It is therefore possible that the effect of treatment is overestimated. We removed this effect by using untreated MMSE values in the NPI, ADL and IADL progression equations for treated patients. The incremental costs and QALYs presented in Table 16 show the same costs but with a reduction in QALYs, as expected.

Table 16. New Basecase without MMSE effect on NPI, ADL and IADL

	New Basecase			New Basecase without MMSE treatment effects carrying over into NPI, ADL and IADL		
	Cost £	QALYs	ICER £	Cost £	QALYs	ICER £
Mild	-3563	0.164	-21713	-3563	0.136	-26130
Moderate	-1763	0.111	-15824	-1763	0.093	-19001

4.2.2. Combined effect of the exploratory studies so far

This section presents the results of the new basecase model after making several changes to the assumptions to explore the combined effect. These were a) fixing the proportion of patients institutionalised across the severity levels; b) including the impact of institutionalisation on caregiver utility and c) removing the MMSE treatment effect from the

NPI, ADL and IADL progression equations. Essentially, this study is the cumulative effect of all the exploratory studies performed so far and the results are shown in Table 17 below.

Table 17. New Basecase with combined exploratory analysis

	New Basecase			New Basecase: combined exploratory analysis		
	Cost £	**QALYs**	**ICER £**	**Cost £**	**QALYs**	**ICER £**
Mild	-3563	0.164	-21713	2186	0.085	£25831
Moderate	-1763	0.111	-15824	1,826	0.058	£31389

4.2.3. Regular update interval

As described in section 3.1.2, the patient's disease status is updated regularly every 3 months and at these time points the costs and QALYs accrued since the last update are calculated. Updates are also made when other events occur such as stopping treatment or death but the timing of these events are unique to each patient. The new basecase model was run for different update intervals and the results are presented in Table 18. There seems to be a clear pattern, as the update period increases the cost savings and QALY benefits decrease and vice versa. We cannot be sure why the costs and QALYs vary in a systematic way in relation to the time period between updates. It may be due to the fact that the patient's attributes at the end of the period are applied to the whole period since the last update without any sort of half cycle correction being used to reflect the fact that their attributes have been changing over that time period. If the patient's utility is falling over time, this would systematically underestimate the QALYs accrued by the patient. If the patient's utility is falling faster in the untreated arm than in the treated arm, one would expect this error to overestimate the QALYs gained by treatment more for less frequent updates. Whilst here we see that the QALY gains are greater for more frequent updates. The cause of this behaviour has not been identified during our examination of the model and therefore we cannot exclude the possibility that it may be due to an error in the model logic.

Table 18. New Basecase with different update intervals

	Mild Population		Moderate Population	
Update period	**Cost £**	**QALYs**	**Cost £**	**QALYs**
30 days	-4247	0.184	-2172	0.126
60 days	-3600	0.166	-1784	0.113
NewBasecase (90 days)	-3563	0.164	-1763	0.111
120 days	-2942	0.149	-1481	0.102

4.2.4. Distribution of sampled life-expectancy estimates

We generated samples of 5000 patients using the distributions that are applied in the model (using Microsoft Office Excel 2007) and compared the summary parameters for these with the MRC CFAS data. The median and interquartile range for each are presented in Table 19 below. The median survival estimates appear to match closely at older ages, but there are differences of up to 0.8 years in some age categories between the trial data and the sampled population which is being used to represent the distribution observed in the trial.

Table 19. Median (and interquartile) survival estimates

	MRC CFAS study (Table 10 of Appendix H)		**5000 patients sampled from the distribution used in the model**	
Age	**Women**	**Men**	**Women**	**Men**
65 to 69	7.5 (4.8-NA)	NA (9.1-NA)	8.1 (5.5 – 10.0)	11.8 (9.1 – 11.3)
70 to 79	5.8 (3.6-8.3)	4.6 (3.0-8.6)	6.0 (3.6 – 8.1)	5.4 (2.9 – 7.8)
80 to 89	4.4 (2.8-7.0)	3.7 (2.5-6.3)	4.8 (2.7 – 6.6)	4.2 (2.4 – 5.8)
>=90	3.9 (2.4-5.2)	3.4 (1.5-5.5)	3.9 (2.4 – 5.2)	3.4 (1.5 – 5.5)

We investigated the sensitivity of the model to differences in the survival estimates by running the model with the survival times fixed at the median and interquartile values taken from the MRC CFAS study.

Table 20. New Base case and new base case with survival fixed at median, upper and lower interquartile range

	Cost £	QALYs	ICER £
Mild			
New base case	-3,563	0.164	-21,713
New base case with survival fixed at median survival	-3,857	0.180	-21,395
New base case with survival fixed at lower IQR	-2,669	0.129	-20,631
New base case with survival fixed at upper IQR	-4,721	0.214	-22,102
Moderate			
New base case	-1763	0.111	-15824
New base case with survival fixed at median survival	-2,085	0.127	-16475
New base case with survival fixed at lower IQR	-1,580	0.105	-15,056
New base case with survival fixed at upper IQR	-2,239	0.133	-16,880

For males aged <70, no median or upper IQR are provided so we assumed that the width of the IQR from males ages 70 to 80 could be applied to estimate the median and upper IQR as 10.7 and 14.7 respectively. For females aged <70 no upper IQR is provided so we assumed that the width of the IQR from females ages 70 to 80 could be applied to estimate the upper IQR as 10. These results show that whilst the cost-effectiveness estimate is sensitive to changes in the survival inputs, treatment still dominates no treatment even when applying the lower IQR for survival from the MRC CFAS study.

5. CONCLUSIONS

Having examined the model, we have been able to provide a more detail description of how the model functions including the fact that it does not employ a pure DES approach but actually incorporates elements of individual sampling alongside some cohort modelling methods. In particular, it uses a cohort approach to estimate the costs of care and patient utilities based on the probability of institutionalisation rather than sampling the location of care for each patient.

We have found and corrected several errors in the implementation of the model methods within the ARENA model. The corrections did not significantly alter the cost-effectiveness estimates, with donepezil dominating no treatment for both mild and moderate AD**.** However, we have continued concerns that there may be further errors within the model as we found behaviour which we could not explain when examining the use of an alternative update frequency. We also have unresolved concerns regarding the way in which the model samples its population.

We found that the deterministic estimates of the ICER overestimated the cost-effectiveness of donepezil compared to the expected ICER obtained from the PSA analysis. This suggests that a robust PSA analysis is needed to determine the cost-effectiveness of donepezil. However, we also had significant concerns regarding the implementation of the PSA analysis and therefore the PSA results should be treated with caution.

We have conducted several exploratory analyses to establish what the cost-effectiveness results would be if changes were made to some of the more questionable model assumptions. These were the relationship between MMSE and institutionalisation, the impact of institutionalisation on care giver utility and the potential overestimation of treatment effects that may be caused by the inclusion of the MMSE treatment effect within the NPI, ADL and IADL progression equations. Exploratory sensitivity analyses showed that the ICER for donepezil compared to no treatment could be as high as £26,000 per QALY in mild AD and £31,000 per QALY in moderate AD if alternative plausible assumptions are made for each of these key model assumptions.

6. REFERENCES

1. IGLEHART, D. L., "Simulating Stable Stochastic Systems, V: Comparison of Ratio Estimators," Naval Res. Logist. Quart., Vol. 22 (1975), pp. 553-565.