
Monograph:
Brazier, J. (2001) Quality of life evidence for patients with Alzheimer's disease : Use of existing quality of life evidence from the ADENA trials to estimate the utility impact of Exelon® (Appendix 1 - Utility evidence). Discussion Paper. (Unpublished)

HEDS Discussion Paper 01/01

HEDS Discussion Paper 01/01


ScHARR
School of Health and
Related Research

Appendix 1 – Utility evidence

Quality of Life Evidence for patients with Alzheimer's disease

Use of Existing Quality of Life evidence from the ADENA trials to estimate the utility impact of Exelon®

- **John Brazier –**

- **2001 -**

1) Introduction

In this Appendix we utilise the Mini-Mental State Examination (MMSE) score of patients with Alzheimer's Disease to establish a relationship between disease progression and quality of life measures and we also compare our results to findings from the literature review about Alzheimer's patients utility.

2) Use of Existing Quality of Life evidence from the ADENA trials to estimate the utility impact of Exelon®

The purpose of the work reported in this first part of the Appendix is to map the outcome measures used on the ADENA Exelon® (rivastigmine) drug trials onto the Health Utilities Index (HUI III Furley et al 1998) utility index in order to create a Quality of Life measure that can be used in a cost- effectiveness analyses of the drug. The mapping process uses the questionnaires involved in eliciting ADAS- Cog, PDS and CIBIC+ scales and, by a comparison of the questions and multiple- choice answers in those questionnaires with those in the HUI III questionnaire, allows a utility index to be constructed.

The HUI III utility index was chosen, rather than another utility index because it includes Cognition as a dimension on the multi- attribute scale. This makes the assessment of utility scales far simpler because of the large cognitive element in Alzheimer's Disease. The outcome measures used in the ADENA trials therefore would be particularly appropriate for such mapping. This does not occur on other utility indices, while HUI II is not as tightly defined as HUI III. Furthermore HUI III has the advantage of having its weights based on the standard gamble (Neumann 1998)

The HUI III utility index is a general utility scale for health states that has a wide range of application across all types of disease. (Feeny et al. 1996) It is not tailored specifically to Alzheimer's Disease (AD) so that not all the dimensions used in assessing illness will be suitable for describing its symptoms. However, by contrast, the outcome measures used in the ADENA trials *are* well suited for describing and measuring the symptoms of AD. When mapping between the outcome measures and the utility scale, one needs to decide which parts of the utility scale are relevant to AD. In the case of the HUI III index the Speech, Cognition, Ambulation and Emotion dimensions were deemed relevant, while the Hearing, Vision, Dexterity and Pain dimensions were deemed irrelevant to the condition of AD.

Although the latter four dimensions were deemed irrelevant to AD they could not be ignored. HUI III is a multi- attribute utility index, composed of several dimensions that need to be combined together. The function used to combine the dimensions together is multiplicative. This means that the dimensions are not independent of each other, so that some realistic value must be found for each of the irrelevant dimensions. The method used was to put in a value obtained from the average value for each dimension derived from previous work (c.f. Neumann 1999). For each dimension, this average value was rounded to the nearest utility number, corresponding to a health state on that dimension.

The remaining four dimensions were mapped onto the three outcome measures as follows. Speech and Cognition were mapped from ADAS- Cog; Ambulation was mapped from the PDS and Emotion was mapped from the CIBIC+ (Novartis 1997 for source of questionnaires etc.). Not all the mappings were simple and all

required certain assumptions to be made in order for the mapping to be valid. Even the simplest required a certain amount of interpretation of the item responses to make them fit. One aspect of this, for example, was the setting of the questionnaires. HUI III is very general and is meant to refer to the patient's whole life at that point in time. By contrast, the three outcome measures refer to the clinician's impression of the patient in the surgery. An assumption had to be made therefore that this impression applied outside the surgery.

The first dimension to be mapped was that of speech, which was mapped onto "Spoken Language Ability" in ADAS- Cog. The latter is a six point scale which goes from 0, "no instance where it is difficult to understand the patient" to 5, "severe". The mapping here was quite simple since the two end-points correspond across scales quite well and the second to fourth scores in ADAS- cog are similar to the second to fourth scores on HUI III. The only manipulation was that point 1 on ADAS-cog, "very mild", was also counted as 1 on HUI III, as well as point 0.

Table 3. 3. 1 Mapping of Speech items

ADAS-Cog -Speech	HUI III - Speech
0,1	1
2	2
3	3
4	4
5	5

The second dimension to be mapped was that of Cognition. This was by far the most complex of the four dimensions, in that no less than five items from ADAS- Cog had to be mapped onto the HUI III cognition dimension. Furthermore, the Cognition dimension is not directly derived from the HUI III questionnaire but from two "sub- dimensions"; Thinking and Memory, which are then combined. Of the five items mapped onto cognition, three were mapped onto memory and two onto thinking.

The three items from ADAS- cog mapped onto the Memory sub- dimension were Word Recall, Word Recognition (which involved a memory task) and Remember Instructions. First of all each individual item was mapped onto Memory. In the case of Word Recall and Word Recognition this created a problem because, in both cases, the scales were not discrete. This meant that the memory scale (a four point scale) had to be mapped to intervals in the word recall scale that varied from a low of 0 to a maximum of 10. Similarly, the memory scale mapped to intervals in the word recognition item. Remember Instructions, by contrast, was a discrete six-point scale that had to be mapped to the four- point scale of Memory.

Table 3. 3. 2 Mapping of items onto Memory

Word Recall		Word Recognition		Remember Instructions	
HUI III	**ADAS- Cog**	**HUI III**	**ADAS -Cog**	**HUI III**	**ADAS- cog**
0-2.5	1	0-2.5	1	1	0
0.25-4.5	2	2.5-5.5	2	2	1,2
4.5- 9.5	3	5.5-11.5	3	3	3,4
9.5- 10	4	11.5-12	4	4	4

For the Thinking sub- dimension of the Cognition dimension, the two items mapped from ADAS- Cog were Comprehension and Ideational Praxis. The former evaluated the patient's ability to understand speech, while the latter evaluated the patient's ability to follow complex instructions. Both were discrete, so there were no problems with continuity. Ideational Praxis was relatively simple to map, with the two endpoints similar. However, Comprehension was more problematic in that the lower endpoints did not match. In particular, it was not thought that "severe" in Comprehension matched "unable to think or solve any problems" which suggested a far worse problem. This meant that the lowest point in Comprehension only mapped onto the second lowest point on Thinking.

Table 3. 3. 3 Mapping of items onto thinking

Ideational Praxis		Comprehension	
HUI III	**ADAS- Cog**	**HUI III**	**ADAS- Cog**
1	0	1	0,1
2	1	2	2
3	2,3	3	3
4	4	4	4,5
5	5	5	***

Having mapped the individual items in ADAS- cog onto the Thinking and Memory sub- dimensions on HUI III, it was necessary to combine the mappings together to produce scores for the sub- dimensions. This was done, in each case, by taking a mean of the mappings and rounding to the nearest whole number. This gave the mapped Thinking and Memory sub- dimensions.

Getting from the sub- dimensions to the Cognition dimension was a complex process in itself. In HUI III the Cognition dimension is formed by assigning numbers to descriptions formed from combinations of the thinking and memory sub- dimension answers (Furlong et al. 1998). However, out of a possible twenty combinations of sentences possible by combining the Thinking and Memory sub- dimensions together, only six have been labelled with a number in the cognition dimension. Unfortunately the six labelled do not cover all the

combinations which exist in the data. The solution to this problem was to make assumptions about which of these combinations could be counted as belonging to which number.

Table 3. 3. 4 Combining Memory and Thinking into Cognition

		THINKING				
		1	2	3	4	5
MEMORY	1	**i**	**ii**	*iv*	*v*	*vi*
	2	**iii**	**iv**	*iv*	*v*	*vi*
	3	*v*	*v*	*v*	**v**	*vi*
	4	*vi*	*vi*	*vi*	*vi*	**vi**

In the table above, the roman numerals give the ranking on the cognition scale of combinations of the memory and thinking sub- scales. The numerals in bold are those recommended in the HUI III questionnaire. The numerals in italics were decided by the assumption that if the cell shared a point on either of the subscales with one of the bold numerals, which was the lowest possible, then it would have the same point on the cognitive scale as that numeral.

The third relevant dimension on HUI III was that of Ambulation. This was mapped onto Walking Ability on the PDS scale. This presented three problems, one technical and two with interpretation. The latter is related to the fact that the link between the two scales is only partial. The ambulation scale is concerned with the ability to walk safely, but is also concerned by the need for walking aids. By contrast, the walking ability scale, while also being concerned with walking safety, is in addition concerned with the patient's propensity to become lost. The mapping between the two is therefore a partial mapping between the ability to walk safely on both scales. One result of this partial relationship is that the lowest points on the scale do not match. On HUI III this is "Cannot walk at all" whereas the PDS merely describes the patient as being "unsafe" to leave the house. It was decided that the latter corresponded better with point 5 on the Ambulation Scale.

The technical problem was that the PDS scale is measured on a Visual Analogue Scale rather than on a discrete scale such as the ADAS- cog measure. This was solved by assuming that the scale was equally split up between the points used on the Ambulation scale. Point 1 on the ambulation scale, for example, was assumed to match up with 100 to >80 on the walking ability scale. Point 5 by contrast corresponded to a range of 0 to 20.

Table 3. 3. 5 Mapping onto Ambulation

PDS- Walking Ability	HUI III -AMBULATION
>80	1

60-80	2
40-60	3
20-40	4
0-20	5

The fourth relevant dimension on the HUI III scale is that of Emotion, which principally measures the amount of depression experienced by patients. This was fairly easily mapped onto the depression subscale of the patient's interview (attitude/behavioural section) on the CIBIC+ measure. The depression subscale goes from 0 (not present) to 3 (Present with emotional and physical concomitants. The HUI III index goes from 1 to 5. However 5 on the scale (So unhappy that life is not worthwhile) seemed substantially worse than 3 on the CIBIC+ scale, so 4 is mapped onto 3 instead. The mapping is shown in the table below:

Table 3. 3. 6 Mapping onto Emotion

CIBIC+- Depression	**HUI III- Emotion**
0	1,2
1	3
2,3	4

Each dimension was then allocated a pre- determined utility score depending on where the dimension was mapped on the questionnaire. Given the four relevant dimensions and the four irrelevant dimensions in the HUI III scale, it was possible to construct a full utility index using a multiplicative function to combine the utility elements together. The formula used was:

$$u = 1- (1/c)(\Pi_j (1+c * c_j * (1-u_j))-1)$$

where u is the total utility, u_j is the utility for an dimension in the utility function, c_j is the constant for each dimension and c is a universal constant. The constants had been estimated from previous work (see Furlong et al. 1998).

Once this was done then the averages of the utility scores between baseline and 26 weeks was calculated for both Exelon and placebo treatments. The utility scores for intermediate points were ignored because these points did not appear consistently in the data. Instead it is assumed that the change in utility between the two points is linear. The averages were then used to calculate the gain/ loss in utility over time. The tables below give the utility averages for placebo and rivastigmine patients at baseline and 26 weeks. They also give the differences between baseline and 26 weeks, representing the improvement or decline of the patients in that

time. Also calculated is the difference between the differences- or how much of an improvement rivastigmine is over placebo. The two tables are for all doses and for high dose only.

Table 3. 3. 7 Mean Utility Results from Pooled Trials (303 and 352) – All doses

	Baseline	**26 Week**	**Difference**	**95% Confidence Interval**
Placebo	0.703425	0.678378	0.025047	±0001379
Exelon®	0.682718	0.690545	-0.00783	±0.005315
		Difference	0.032875	

Table 3. 3. 8 Utility Results from Pooled Trials (303 and 352) -High Dose Rivastigmine Only

	Baseline	**26 Week**	**Difference**	**95% Confidence Interval**
Placebo	0.703425	0.678378	0.025047	±0.001379
Exelon®	0.670813	0.702671	-0.03186	±0.000931
		Difference	0.056906	

As can be seen from the tables above, in the first table the improvement as a result of using rivastigmine is less than that assured in the Wessex DEC report (1998), whereas in the second table the numbers are higher. This is in line with what would be expected from the overall results where only the high- dose rivastigmine has a significant effect on the progress of the disease.

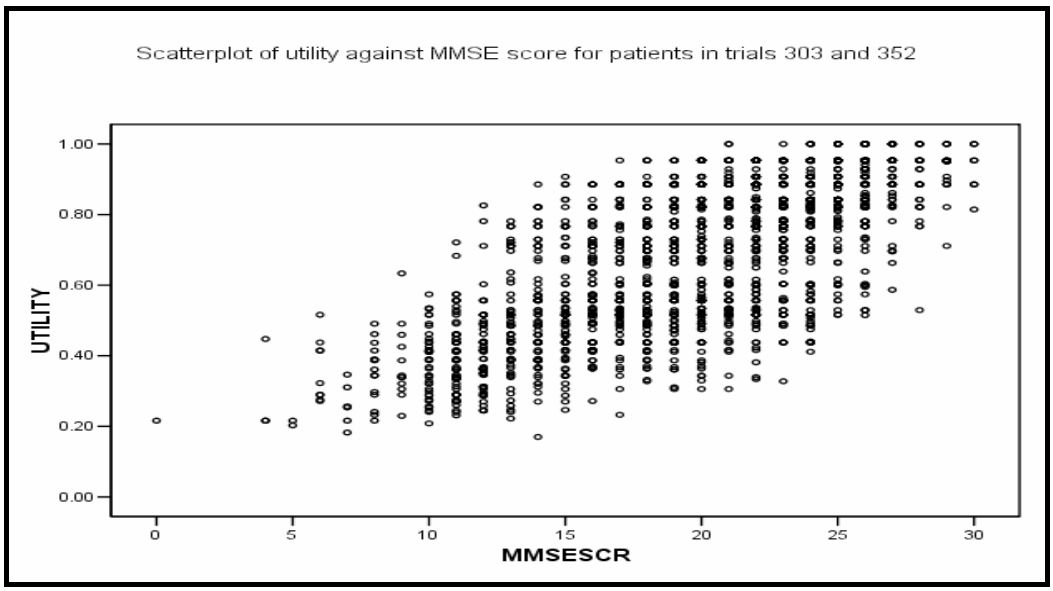
In order to test for the robustness of this mapping, a similar process of averaging the scores for each individual item on the ADAS- cog, CIBIC+ and PDS scales was used. This averaged over all the cases representing baseline and 26 weeks for each item used in these measures. The differences in performance on each of these scales were calculated. The results were then normalised so that the item averages varied between 0 and 1 to allow comparison with the utility scores. In absolute terms, the magnitude of the score changes varied between 0.06823 (remember instructions) and 0.002964 (Emotion). These scores are within the range expected, given the high dose utility difference in performance (0.056906) and the all- dose utility difference in performance (0.032875).

After having constructed utility scores for each patients the utility scores and their corresponding MMSE scores were put into a regression model.

3) Summary of ADENA evidence

The figure below shows the relationship of utility to MMSE score on the two ADENA trials (303 and 352) population. Patients' baseline and 26-week data were used.

Figure 3. 3. 1. Utility against MMSE score for patients in the ADENA trials



A trend of higher utility related to higher MMSE status can be seen and statistical regression confirmed the slope and significance of the relationships.

Table 3. 3. 9. Result of the statistical regression of individuals' MMSE scores versus utility

Variable	B	95% Confidence Interval for B		t-value	Significance
Constant	0.0982	0.0735	0.1228	7.8130	0.0000
MMSE score	0.0298	0.0286	0.0310	48.4330	0.0000

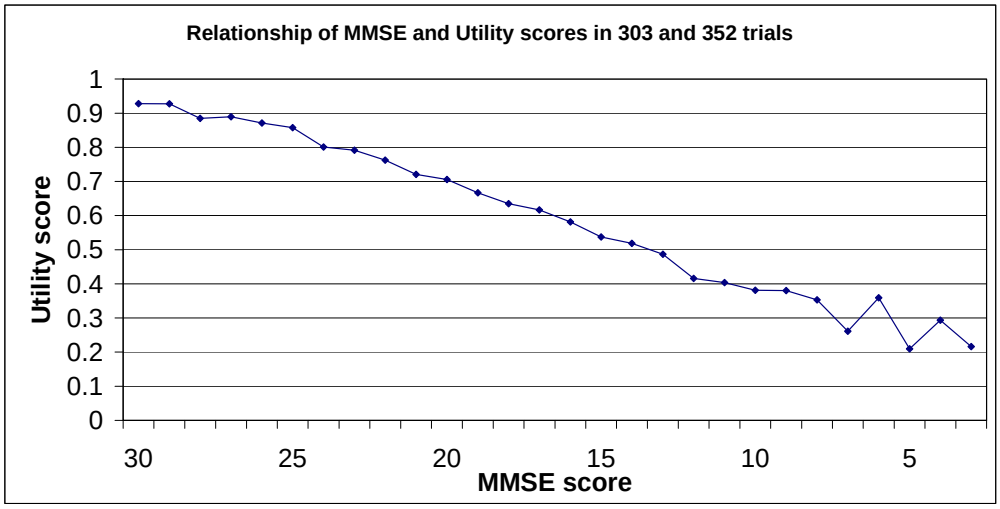
Regression type	Dependent Variable	R Square	Adjusted R Square	F Change	Significance of F Change
linear	UTILITY	0.5185	0.5183	2345.7948	0.0000

The conclusion is that a reduction in MMSE of 1 point is roughly equivalent to a reduction in utility of 0.03.

These results match reasonably well with the quality of life results shown in Table A 3.3.7 and Table A 3.3.8

Figure A 3.3.2 below shows the mean utility for each MMSE score from the base line and the 26 week data. It can be clearly seen that the reductions in MMSE correspond well with reductions in utility.

Figure 3. 3. 2 Relationship between utility and MMSE scores in the ADENA trials



Number of observations	7	21	26	50	153	166	165	173	178	138	161	135	145	104	107	91	78	65	64	57	48	12	15	6	9	2
MMSE SCORE	30	29	28	27	26	25	24	23	22	21	20	19	18	17	16	15	14	13	12	11	10	9	8	7	6	5
Mean utilty score	0.93	0.93	0.88	0.89	0.87	0.86	0.80	0.79	0.76	0.72	0.71	0.67	0.64	0.62	0.58	0.54	0.52	0.49	0.42	0.40	0.38	0.38	0.35	0.26	0.36	0.21

At 6 months rivastigmine clearly gives a quality of life benefit. The quality of life benefit is proportional to the delay in MMSE reductions.

4) Literature Evidence on Quality of Life with patients with Alzheimer's disease

There is very little literature on the relationship between dementia and QALYs. However there is a short chapter by Peter Neumann, Richard Herman, and Milton Veinstein entitled "Measuring QALYs in Dementia" which is contained within the book Health Economics of Dementia (edited by Anders Wimo, Bengt Jonsson, Goran Karlsson and Bengt Winblad). The researchers review the various different methodological issues in relation to developing QALYs.

The authors then move on to undertake an analysis of Quality of Life data which used "A companion, cross sectional study of 528 care givers of Alzheimer's Disease patients, stratified by disease stage (mild, moderate, and severe based on the clinical dementia rating staging), and setting of care (community and nursing home) to obtain the utility weights. The 528 patients included 201 mild, 175 moderate and142 severe; of these 354 were cared for in community settings and 164 in nursing homes."

"In the cross sectional study, the study investigators administered the health utilities index mark 2. This was chosen because it provides a means of obtaining community based preference rates in accordance with the US panel on cost effectiveness recommendations for reference case analyses and because weights are based on the standard gamble method. Another advantage of the HUI 2 is that, unlike other preference weighted health status classification systems, the HUI 2 contains cognition as a separate attribute which may make it more sensitive to changes in Alzheimer's Disease stage. The HUI 2 questionnaire was completed by primary care givers as proxy respondents. The responses were converted into preference weights using the HUI 2 multi-attribute utility function (Torrance et al, 1995)."

Table A 3.3.11. shows the weights obtained from the cross sectional study.

Table 3. 3. 10. Estimates of quality-of-life weights across Alzheimer's disease stages and settings of care

Stage/Setting	Quality-of-Life Weights	
	Patients	Caregivers
Mild AD		
Community	0.68	0.86
Nursing Home	0.71	0.86
Moderate AD		
Community	0.54	0.86
Nursing Home	0.48	0.88
Severe AD		
Community	0.37	0.86
Nursing Home	0.31	0.88

Source: Neumann et al (1998) Table 3.4.5

In another cross sectional study Neumann et al (1999) assessed the utility scores of 679 Alzheimer's disease patient/caregiver pairs in different disease stages. Patients' AD stage was determined by clinicians using Clinical Dementia Rating (CDR) scale, which closely corresponds to MMSE. They classified AD into one of six categories. HUI 2 scores were converted into global utility score with using the multi- attributable utility function. See results of the Neumann et al (1999) study.

Table 3. 3. 11 Patients HUI score by CDR severity level and service setting

AD severity	Patients	Caregivers
Questionable (CDR=0.5)	0.73	0.88
Mild (CDR=1)	0.69	0.87
Moderate (CDR=2)	0.53	0.87
Severe (CDR=3)	0.38	0.86
Profound (CDR = 4)	0.27	0.9
Terminal (CDR = 5)	0.14	0.93

Source: Neumann et al (1999) Table 3.4.6

Leon et al (2000) analysed the same patient population but put more emphasis on service utilization. They stratified the same set of patients into only 3 CDR categories and used HUI scores to measure utility. This approach gives less information but it is more comparable to the ScHARR results. Patients utility scores were in mild, moderate and severe categories: 0.70, 0.53, 0.34 respectively. These scores seem to highly correspond to both of Neumann's previous analyses.

Table 3. 3. 12. Patients HUI score by CDR severity level and service setting

Disease severity	Service setting		HUI scores	HUI scores average
Mild AD	Community	AMCs	0.69	

		MCOs	0.67	**0.7**
	Residential	Assisted Living	0.74	
		Nursing Home	0.71	
Moderate AD	Community	AMCs	0.53	**0.53**
		MCOs	0.56	
	Residential	Assisted Living	0.56	
		Nursing Home	0.48	
Severe AD	Community	AMCs	0.36	**0.34**
		MCOs	0.38	
	Residential	Assisted Living	0.35	
		Nursing Home	0.31	

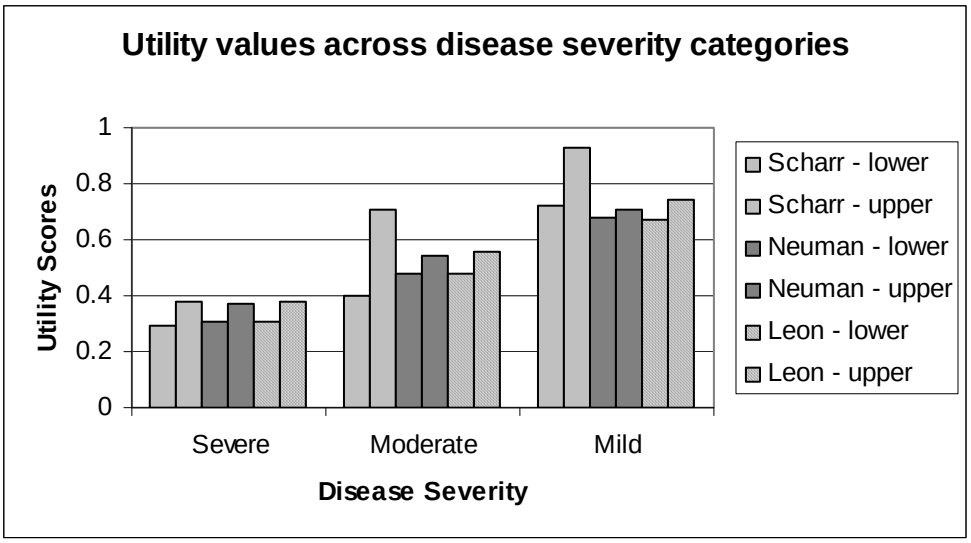
Source: Leon et al (2000)

Relationship between literature's findings and ScHARR analysis

Comparing the results of the literature review with those set out in first part of this Appendix it can be seen that if we were to assume broadly that MMSE groups 0-10 related to severe, 11-20 related to moderate, and 21-30 related to mild then the utility for clients in the mild, moderate and severe groups are within very similar ranges.

For example 0.68-0.71 (Neumann et al 1998) or 0.67-0.74 (Leon et al 2000) for mild Alzheimer's compares with a range of 0.72 up to 0.93 for MMSE groups 21-30, whilst 0.48-0.54 or 0.48-0.56 for moderate Alzheimer's disease compares with a range of 0.40 to 0.71 for MMSE 11-20 and 0.31-0.37 or 0.31-0.38 for severe Alzheimer's compares with a range of 0.29 to 0.38 for MMSE scores 4-10.

Figure 3. 3. 3 Comparison of ScHARR results and US evidence on disease severity and utility



An important issue is that we could not consider any quality of life change following institutionalisation. The other studies' use of utility do separate QoL for institutionalised and community settings. Therefore, our estimates are possibly conservative.

5) Conclusions

1. The Neumann et al. (1998, 1999) and Leon et al (2000) methodology for assessing quality of life in Alzheimer's cases gives broadly similar results to our own analysis of the relationship between MMSE and utility.

2. The Neumann et al. (1998, 1999) and Leon et al (2000) methodology utilising the HUI 2 and the Torrance multi-attribute utility function is slightly different to our own methodology set out in the first part of this Appendix

3. However using a slightly different methodology and a different base data set still provides effectively the same results for quality of life weights for Alzheimer's disease stages.

4. The US results do not give any indication of invalidating our own analysis, indeed they support the broad conclusions. Therefore we can use for the calculation of utility in our cost effectiveness model the following equation derived from the regression in Table 3.3.9:

$$\text{Utility} = 0.0982 + \text{MMSE} * 0.0298$$

Bibliography

1) Feeny D.H., Torrance G.W. and Furlong W.J. “Health Utilities Index” Ch.26 in “Quality of Life in Clinical Trials (1996) 2nd Edition, edited by B.Spilker Lippincott- Raven Publishers, Philadelphia

2) Furlong W., Feeny,D. et al (1998) “Multiplicative Multi- attribute Utility Function for the Health Utilities Index Mark 3 (HUI 3) System: A Technical Report” McMaster University Centre for Health Economics and Policy Analysis Working Paper 98-11

3) Novartis- (1997) “Integrated Summary of Effectiveness” Project SDZ ENA 713

4) Neumann P.J., Hermann R.C. and Weinstein M.C. (1998) “Measuring QALYs in Dementia” ch. 3.3 in “Health Economics of Dementia” edited by Wimo A., Jonsson B., Karlsson G. and Winblad B. John Wiley and Sons Ltd.

5) Neumann P.J. Sandberg E.A. et al (1999) “A Comparison of HUI 2 and HUI 3 Utility Scores in Alzheimer’s Disease” Working Paper

6) Wessex Institute (1998) “Rivastigmine (Exelon) in the treatment of semile dementia of the Alzheimer type (SDAT)” Development and Evaluation Committee Report no.89

7) Neumann PJ and et al. A comparison of HU12 and HU13 utitlity scores in Alzheimer's Disease. *Medical Decision Making* 20:413-422, 2000.

8) Neumann PJ, Kuntz KM, Leon J, Araki SS, and et al. Health utilities in Alzheimer's Disease: A cross-sectional study of patients and caregivers. *Medical Care* 37 (1):27-32, 1999.

9) Leon J.Neumann PJ.Hermann RC.Hsu M.Cummings JL.Doraiswamy PM.Marin D. Health-related quality-of-life and service utilization in Alzheimer's disease: a cross-sectional study. *American Journal of Alzheimer's Disease & Other Dementias* 15 (2):94-108, 2000.