Comparing EQ-5D scores for comorbid health conditions estimated using five different methods

Ara Roberta, Brazier John

Health Economics and Decision Science, ScHARR, The University of Sheffield, 30 Regent Street, Sheffield S1 4DA. UK

Telephone: 44114 222 0788
Fax 44 114 272 4095
Email r.m.ara@sheffield.ac.uk

Acknowledgements
The Health Survey for England is commissioned by the Department of Health and conducted by the Joint Health Survey Unit of National Centre for Social Research and Department of Epidemiology and Public Health at University College London. Ethical approval for the Health Survey for England was obtained from the London Multi-Centre Research Ethics Committee.

Keywords: health state utility values, comorbidities, quality of life, EQ-5D

ABSTRACT

BACKGROUND: While health state utility values (HSUVs) for many single health conditions are now in the public domain, due to the large number of possible combinations of comorbid health conditions (CHC) the HSUVs for these are not readily available. As a consequence, HSUVs for CHCs are frequently estimated using data obtained from cohorts with single health conditions. With researchers presenting conflicting results, there is currently no consensus on the most appropriate method to estimate HSUVs for CHCs.

OBJECTIVE: The objective of the study was to assess the accuracy of five different methods in the same dataset.

METHODS: EQ-5D data (n=41,174) from the Health Survey for England was used to compare HSUVs generated using the following techniques: the additive, multiplicative and minimum methods, the adjusted decrement estimator (ADE), and a linear regression model.

RESULTS: The additive and multiplicative methods under estimated the majority of HSUVs and the magnitude of the errors increased as the actual HSUV increased. Conversely, the minimum and DAE methods over estimated the majority of HSUVs and the magnitude of errors increased as the actual HSUV decreased. Although the simple linear model produced more accurate results than the others, there was a tendency to under predict higher HSUVs and over predict lower HSUVs and 20% of the errors were greater than the MID (|0.074|) for the EQ-5D. We found the magnitude and direction of mean errors in the estimated scores could be driven by the actual scores being estimated in addition to the technique used and in general the HSUVs estimated using an adjusted baseline were more accurate.

We found the additive and minimum methods performed very poorly in our data. While the simple linear model gave the most accurate results, the model requires validating in external data and additional research exploring alternative model specification is warranted. Our comparison of errors in subgroups of actual EQ-5D scores highlights the need to present additional data when reporting results of analyses in this area as conclusions using average errors in truncated ranges could be misleading.

BACKGROUND

Throughout the world economic evaluations in health care are used by policy decision makers to make informed decisions on whether new treatments should be reimbursed. Clinical effects are measured in terms of health related quality of life (HRQoL) impacts and costs include the resources used and the intervention costs. Treatments are appraised using a decision rule based on an incremental cost effectiveness ratio which utilises a generic HRQoL measure to quantify the cost per quality adjusted life year (QALY).[1] The QALY is a metric that combines the duration of time spent in a health state with the HRQoL associated with the health state i.e. the health state utility value (HSUV).[2] The rationale for the use of the QALY is to facilitate consistent reimbursement recommendations across all disease areas and systems thus allowing the optimal allocation of resources. However, the use of inconsistent methods can undermine the aim for consistent decision making.

Decision analytic models used to generate the cost effectiveness ratios are mathematical models which represent the health condition or system under appraisal. Due to the ageing population and the increasing prevalence of concurrent health conditions, these models frequently describe comorbid health conditions (CHCs) such as arthritis and stroke.[3] While there are a number of catalogues which provide HSUVs for cohorts with a single health condition, due to the enormous number of possible combinations of CHCs, the HSUVs for these are not readily available. As a consequence, analysts estimate HSUVs for CHCs using data obtained from cohorts with single health conditions. For example, the HSUV for a cohort with the CHC arthritis plus stroke would be estimated using the mean HSUV obtained from a cohort who have a history of arthritis (but not stroke) and the mean HSUV obtained from a cohort who have a history of stroke (but not arthritis). HSUVs are estimated on a cohort level as opposed to an individual patient level as by definition individuals cannot have just a single health condition and a CHC.

The three most commonly cited techniques used to estimate HSUVs for CHCs are the additive, multiplicative and minimum methods. The additive and multiplicative methods assume constant absolute and proportional decrements respectively while the decrement associated with the minimum method can vary depending on the baseline used. Two alternative techniques have recently been suggested. First, the adjusted decrement estimator (ADE), a non parametric estimator based on the minimum method has been proposed.[4] Second, a hybrid model which incorporates terms representing the three traditional methods has been explored using ordinary least square regressions.[5]

The existing evidence base in this area is relatively small and there is currently no consensus on the most appropriate method. Comparison of findings reported in the literature is hindered due to differences in study designs such as the preference based measure estimated, differences in datasets, methodologies compared, ranges of HSUVs estimated, and the statistics used to compare the estimated values.[6-8] The methods can produce very different results and when applied in economic models the method used could potentially influence a policy decision based on a cost per QALY threshold.[9]

The primary objective of the current study is to add to the existing evidence base through comparing all five methods in a single dataset. We use EQ-5D data from the Health Survey for England to examine the results generated using the different techniques. The next section provides a description of the alternative methods followed by the results of our analyses.

METHODS

Methods used to estimate HSUVS: Given two health conditions, condition A and condition B, there are four combinations possible: individuals have condition A but not condition B, individuals have condition B but not condition A, individuals have condition A and condition B; individuals do not have either condition A or condition B. The HSUVs associated with these alternatives are defined as: U_A, U_B, $U_{A,B}$, and $U_{nA,nB}$.

The additive method assumes a constant absolute decrement relative to the baseline and the HSUV for the CHC is estimated using:

$$U_{A,B}^{add} = U_{nA,nB} - \left(\left(U_{nA} - U_A\right) + \left(U_{nb} - U_B\right)\right) \quad \text{(Eqn 1)}$$

If a baseline of perfect health is used, the additive method can be estimated using:

$$U_{A,B}^{add} = U_A + U_B - 1 \quad \text{(Eqn 2)}$$

The multiplicative method assumes a constant proportional decrement relative to the baseline and the HSUV for the CHC is estimated using:

$$U_{A,B}^{Mult} = U_{nA,nB} \cdot \left(\frac{U_A}{U_{nA}}\right) \cdot \left(\frac{U_B}{U_{nB}}\right) \quad \text{(Eqn 3)}$$

If a baseline of perfect health is used, the multiplicative method can be estimated using:

$$U_{A,B}^{mult} = U_A \cdot U_B \quad \text{(Eqn 4)}$$

The minimum method assumes the HSUV for the CHC is equivalent to the minimum HSUV for the single health conditions within the CHC. The HSUV for the CHC is estimated using:

$$U_{A,B}^{min} = \min(U_{nA,nB}, U_A, U_B) \quad \text{(Eqn 5)}$$

If a baseline of perfect health is used, the minimum method is estimated using:

$$U_{A,B}^{min} = \min(U_A, U_B) \quad \text{(Eqn 6)}$$

The adjusted decrement estimator (ADE) has recently been proposed as an alternative method.[4] The ADE assumes the estimated HSUV for the CHC has an upper bound equal to the minimum of the HSUVs from the two single health conditions. Assuming a baseline of perfect health, the HSUV for the CHC is estimated using:

$$U_{A,B}^{ADE} = \min(U_A, U_B) - \min(U_A, U_B) \cdot (1 - U_A) \cdot (1 - U_B) \quad \text{(Eqn 7)}$$

Using an adjusted baseline equation 7 can be written as follows:

$$U_{A,B}^{ADE} = \min(U_A, U_B) - \min(U_A, U_B) \cdot (U_{nA} - U_A) \cdot (U_{nb} - U_B) \quad \text{(Eqn 8)}$$

A simple linear model which incorporates terms that represent the additive, multiplicative and minimum methods has recently been proposed.[5] The model, which is based on decision theory, multi-attribute utility functions, and a prospect theory, [10-12] is defined by:

$$U_{A,B}^{comb} = \alpha + \beta_1 \cdot \min((1 - U_A), (1 - U_B)) + \beta_2 \cdot \max((1 - U_A), (1 - U_B)) + \beta_3 \cdot (U_A \cdot U_B) + \varepsilon \quad \text{(Eqn 9)}$$

The beta coefficients are obtained using ordinary least square regressions and ε represents the residual.

Eqn 9 reduces to the additive method when $\beta_0 = 0$, $\beta_1 = 1$, $\beta_2 = 1$ and $\beta_3 = 0$

Eqn 9 reduces to the multiplicative method when $\beta_0 = 0$, $\beta_1 = 1$, $\beta_2 = 1$ and $\beta_3 = -1$

Eqn 9 reduces to the minimum method when $\beta_0 = 0$, $\beta_1 = 1$, $\beta_2 = 0$ and $\beta_3 = 0$

If an adjusted baseline is used as opposed to a baseline of perfect health, he model could be defined by:

$$U_{A,B}^{comb} = \beta_0 + \beta_1 \cdot \min((U_{nA} - U_A), (U_{nb} - U_B)) + \beta_2 \cdot \max((U_{nA} - U_A), (U_{nb} - U_B)) + \beta_3 \cdot \left(U_{nA,nB} \cdot \frac{U_A}{U_{nA}} \cdot \frac{U_B}{U_{nb}} \right) + \varepsilon \quad \text{(Eqn 10)}$$

There is currently no consistency in the baseline used when estimating HSUVs for CHC and researchers have used a baseline of perfect health,[4,8] "purified" data by dividing all HSUVs by the mean HSUV obtained from individuals with none of the health conditions in a

particular dataset,[5,8] or estimated condition specific baselines.[6] Using a baseline of perfect health overestimates the disutility associated with health conditions,[13] and results generated from analyses using a baseline of perfect health are not comparable to those generated using an adjusted baseline.[9] It has been suggested that alleviating a health condition would at best increase the average HSUV to that observed in cohorts without the health condition.[14] Consequently when estimating HSUVs, a baseline of perfect health may not be the most appropriate technique. The ideal baseline would be the HSUV associated with not having a particular health condition. I.e. the baselines for condition A, condition B, or both condition A and condition B, would be obtained from individuals who do not have the respective conditions. Due to the enormous number of possible health conditions the data required to estimate condition specific baselines are not in the public domain. We used age adjusted baseline HSUVs obtained from respondents who indicate they do not have any of the health conditions identified in the dataset used. For comparison with results in the literature, we also generated results using a baseline of perfect health.

Data

The Health Survey for England (HSE) is an annual survey conducted on randomly selected samples of the population living in private households in England.[15-18] The present study pools data from the 2003, 2004, 2005 and 2006 core HSE using the weight variables for the individual level self-administered questionnaire which adjusts for non response. We used the chronic clinical conditions reported in the HSE which included 39 individual health conditions and 15 grouped health conditions.

The EQ-5D is a widely used generic instrument that contains of five attributes of health status including: mobility, self-care, usual activities, pain/discomfort and anxiety/depression. Each of the attributes is measured by a question with three possible responses: no problem, some problem, or severe problem. The combination of possible responses lead to a total of 243 (3^5) possible health states. A sample of these health states were weighted by the UK general public using time trade-off techniques and the resulting algorithm is used to calculate the preference-based HSUVs used in the current study.[19] The preference-based index has a range of -0.59 to 1, whereby 1 represents perfect health, 0 represents death and negative values represent health states considered to be worse than death.

Statistical Analyses

The HSE data were subgrouped ($n \geq 100$) by health conditions using a) groups with comorbid pairs of health conditions (condition A and condition B) regardless of other health conditions, and b) groups with just one of the individual health conditions within each comorbid pair (i.e.

condition A but not condition B; condition B but not condition A). The mean EQ-5D scores (referred to as "actual" EQ-5D scores from here on) and the mean age for each subgroup were calculated. The mean EQ-5D scores from the subgroups with the single health conditions were then used to estimate mean EQ-5D scores (referred to as "estimated" EQ-5D scores from here on) for the corresponding CHC using the methods described previously.

Performance of methods

As we are interested in how well the methods estimate the actual EQ-5D scores, we assess the results in terms of the errors in the estimated HSUVs. In addition to using the mean absolute errors (MAE), and root mean squared errors (RMSE), we examine the proportion of estimated values within the minimum important difference (MID) for the EQ-5D (MID = 0.074)[20] and the magnitude and direction of errors across the EQ-5D range.

RESULTS

The pooled data included 41,174 cases with EQ-5D scores. The individual data covered the full index (-0.5940 to 1) and the mean HSUV for the cohort was 0.8679 (se 0.0014). Approximately 45% (18725/41174) of respondents reported having at least one chronic health condition. The mean HSUV for this subgroup was 0.7565 (se 0.0026) compared with 0.9493 (se 0.0009) for those who reported no chronic health condition. The mean HSUVs for the 97 subgroups ($n \geq 100$) with two health conditions ranged from 0.3596 (se 0.0296) for respondents (n=171) who reported both mental illness/anxiety/depression and arthritis/rheumatism/fibrositis to 0.9165 (se 0.0140) for respondents (n=112) who reported both asthma and hay fever. 69% (67/97) of subgroups had a mean EQ-5D score below 0.6 and none had a negative mean score (Figure 1). As these are mean scores as opposed to HSUVs from individuals, the distribution is relatively normal.

INSERT FIGURE 1: Distribution of mean EQ-5D scores for subgroups (n=97) with two comorbid health conditions

Table 1 provides the results for the linear models obtained using ordinary least square regressions. While the coefficients for the independent variables are all positive, which is as expected given the negative constant terms, none of the coefficients are significant ($p>0.05$). In both models the weight attributed to the maximum disutility are greater than the weight attributed to the minimum disutility and the interaction term has the largest coefficient.

INSERT TABLE 1 : Results from the OLS combination model

Comparing the methods in terms of their accuracy (Table 2), the mean EQ-5D scores of the individual values predicted using the OLS combination models are the closest to the actual mean values (0.5682) when using a baseline of perfect health (mean = 0.5669) and when using an age-adjusted baseline (mean = 0.5671). However, the ranges of the predicted values are smaller than the actual range (actual range = 0.5570). The minimum (ADE) methods produce the smallest range in estimated values at 0.2047 (0.2759) when using a baseline of perfect health, and 0.2715 (0.2415) when using an age-adjusted baseline respectively. While the additive method produces the widest range in estimated values (0.4797 when using a baseline of perfect health and 0.4614 when using an age adjusted baseline), the increased range in estimated values is associated with the largest errors in the individual estimations (0.3320 when using a baseline of perfect health and 0.2792 when using an age adjusted baseline).

The OLS models produce the smallest MAEs and RMSEs in the predicted values when using either a baseline of perfect health or when using an age-adjusted baseline (MAE: 0.047, RMSE: 0.060 for both analyses). Conversely, the additive method produces the largest MAE and RMSE when using a baseline of perfect health (MAE: 0.1411, RMSE: 0.1529). The errors for the additive method are reduced somewhat when using an age-adjusted baseline (MAE: 0.0872; RMSE: 0.1012). While the mean errors give an indication of average accuracy across the full range of estimated values, these statistics do not reveal accuracy in individual estimated or predicted values and there are some substantial errors in the individual estimated HSUVs. The additive method produces the largest individual error for both sets of analyses and the OLS model produces the smallest individual error for both sets.

Of the non parametric methods, the additive and the minimum methods are the least accurate in terms of the proportion of individual values estimated to within a given magnitude of error irrespective of the baseline used. Comparing the multiplicative and the ADE results, the baseline influences accuracy and when using an age adjusted baseline the multiplicative method produces the largest proportion (72% vs 52%) of individual values within the MID for the EQ-5D (|0.074| and the largest proportion (56% vs 35%) accurate to within |0.05| of the actual HSUVs. Conversely, when using a baseline of perfect health, the ADE method produces the largest proportion of values within these measures.

INSERT TABLE 2: Comparing the accuracy of HSUVs generated using the five methods

When plotting the actual and estimated/predicted HSUVs (Figure 2) it can be seen that the values estimated using an age-adjusted baseline are more accurate than those estimated using a baseline of perfect health when using the additive and multiplicative methods. However, the baseline is not as important for the other three methods. For the additive method almost all values are underestimated across the full range of estimated values. For the multiplicative method there is a tendency for the errors in the estimated values to decrease for lower HSUVs with the largest errors in values above 0.6. Conversely, the errors in the minimum and ADE methods increase as the actual HSUV decreases with larger errors observed in estimates for lower HSUVs. Although the errors in the HSUVS predicted using the OLS models are smaller than those in the other methods, there is a tendency to under predict higher HSUVs and over-predict lower HSUVs.

INSERT FIGURE 2: Plot of actual and estimated EQ-5D scores

The data were subgrouped into sets of equal numbers (Table 3) based on the actual EQ-5D score and the mean errors in each of the four groups were examined. Presenting the data in this way reveals additional detail relating to the accuracy of the methods. Using a baseline of perfect health, the additive and multiplicative methods produce smaller errors in HSUVs below 0.56 compared to errors in HSUVs above this. Conversely, the minimum and ADE methods produce smaller errors in HSUVs above 0.562 compared to errors in HSUVs below this. This trend holds regardless of the baseline used. With the exception of the values predicted for the subgroup at the top of the range (EQ-5D greater than 0.624), the OLS model tends to be more accurate than all the non parametric methods although the multiplicative method is the most accurate for the subgroup at the lower end of the range (EQ-5D smaller than 0.514).

DISCUSSION

The purpose of this study was to assess the performance of the three techniques conventionally used to estimate mean HSUVs for comorbid conditions. I.e. the additive,

multiplicative and minimum methods, and the two more recently suggested alternatives, the ADE method and a simple linear regression model. We found the additive and multiplicative methods under estimated the majority of HSUVs irrespective of the baseline used and the magnitude of the errors increases as the actual HSUV increases. Conversely, the minimum and DAE methods over estimated the majority of HSUVs and the magnitude of errors increases as the actual HSUV decreases. Although the simple linear model produced more accurate results than the non parametric estimators, there was a tendency to under predict higher HSUVs and over predict lower HSUVs. There were also some substantial errors in the individual predicted HSUVs with 20% of errors greater than the MID (|0.074|) for the EQ-5D.

A methodological strength of this study is the relatively large range (0.360 to 0.917, 66% smaller than 0.60) in actual mean HSUVs for the combined health conditions. Flanagan et al. assessed the multiplicative method using data from the Canadian Community Health Survey and reported 66% (185/278) of the mean HUI3 scores for cohorts with two health conditions was mild (greater than 0.80).[7] Similarly, when comparing the multiplicative and additive methods using data from the Medical Expenditure Panel Survey (MEPS) the mean EQ-5D scores for cohorts with two comorbid health conditions were reported to be relatively mild (mean EQ-5D range 0.68 to 0.86).[6] The range in actual HSUVs enabled us to assess performance of the methods across subgroups of the EQ-5D index and we found the magnitude and direction of mean errors in the estimated scores could be driven by the actual scores being estimated in addition to the technique used. This suggests that conclusions based on truncated ranges using average errors could be misleading.

Comparing the three original non parametric methods in terms of average errors and proportions of estimated HSUVs accurate to within a given magnitude, when using a baseline of perfect health, we found the additive method was the least accurate and the multiplicative method was the most accurate. When using an age adjusted baseline, the accuracy for both the additive and multiplicative methods increased and the minimum method was the least accurate while the multiplicative method remained the most accurate. These results do not support those reported by other researchers who found the minimum method gave the most accurate results when comparing the three methods in EQ-5D data obtained from the Medical Expenditure Panel Survey (MEPS).[8] The difference in results are easily explained by comparing the ranges of actual HSUVs estimated. The actual HSUVs for the comorbid health conditions in our data ranged from 0.360 to 0.917 with over two thirds of these below 0.60 while the actual HSUVs in Fu's data covered a much smaller range (approximately 0.62 to 0.90).[4] If we examine the average errors in a similar range (0.624 to 1), then the minimum method is more accurate than the additive and multiplicative methods in our data too.

Hu and Fu have recently proposed the ADE method based on analyses of the MEPS data used in the previous study.[4] Using a baseline of perfect health they found the ADE was more accurate than the three traditional methods. We also found the ADE method outperformed the other three non parametric methods in our data when using a baseline of perfect health, but when using an age adjusted baseline, the multiplicative method outperformed the ADE method. Again, when examining the errors for the data subgrouped by actual EQ-5D score, the ADE method performed less favourably for lower HSUVs.

One would intuitively expect that an additional health condition would have a negative effect on health related quality of life and that mean HSUVs for cohorts with comorbid health conditions would therefore be lower than the mean HSUV for cohorts with any of the single health conditions within the comorbid health condition. However, some inconsistencies in HRQoL measurements are to be expected and in our dataset a small proportion (6/97) of the mean HSUVs for cohorts with a comorbid health condition were greater than one of the mean HSUVs for the corresponding single health conditions. It is clear from charts presented in Hu and Fu's article that this anomaly is observed in a substantial proportion of their data as approximately 25% of HSUVs estimated using the minimum method are smaller than the actual HSUV.[4] This is possibly due to the fact that the health conditions in their data have a relatively small affect on HRQoL data and this may contribute to the difference in findings in their dataset.

Although the simple linear model produced more accurate results than the non parametric estimators in our data, none of the coefficients in the model were significant and the model requires validating in external data. The trend to under estimate higher HSUVs and over estimate lower HSUVs suggests that a different model specification may be warranted and additional research exploring alternatives would be beneficial. A limitation of using regressions to explore relationships between HSUVs is that models are unlikely to be valid for HSUVs obtained using different preference-based measures thus each measure would require an individualised model.

One additional problem of estimating HSUVs using the non parametric methods is associated with negative HSUVs. The multiplicative method will produce a positive value when combining two negative values and the additive method could produce estimates below the minimum value on the preference-based index. However, while analysts need to be aware of the potential problem, when estimating mean scores for cohorts, the issues associated with negative scores will arise infrequently.

CONCLUSION

This study makes an important contribution to the evidence in this area as it is the first to compare the five different techniques within the same study. We found the additive and minimum methods performed very poorly in our data. While the simple linear model gave the most accurate results, the model requires validating in external data and additional research exploring alternative model specification is warranted. Our comparison of errors in subgroups of EQ-5D scores highlights the need to present additional data when reporting results of analyses in this area as average errors may not give an accurate picture of overall accuracy.

REFERENCES

1. NICE Guide to the methods of technology appraisal. 2008. London, NICE

2. Gold MR, Siegel JE, Cost-Effectiveness in Health and Medicine. Oxford University Press ISBN 0-19-510824-8

3. Bansback N, Ara R, Ward S, Anis A, Choi HK. Statin therapy in rheumatoid arthritis: a cost-effectiveness and value-of-information analysis. Pharmacoeconomics. 2009;27(1):25-37. doi: 10.2165/00019053-200927010-00004.

Ref catalogiues

4. Hu B, Fu AF, Predicting utility for joint health states, a general framework and a new non parametric estimator. 2010 MDM in press

5. Basu A, Dale W, Elstein A, Meltzer D. A linear index doe predicting joint health state utilities from single health state utilities. Health Economics 2009;18:403-19

6. Janssen M, Bonsel G Estimating preference weights for chronic multimorbidity: Don't add, multiply. Proceedings of the EuroQol Group 2008.

7. Flanagan W, McIntosh C, Le Petit C, Berthelot J. Deriving utility scores for co-morbid conditions: a test of the multiplicative model for combining individual condition scores. Population Health Metrics 2006:4(13) doi:10.1186/1478-7954-4-13.

8. Fu A, Katan M. Utilities should not be multiplied. Medical Care 2008;46:984-90.

9. Ara R, Brazier J, The issues involved in populating an economic model with utility values: moving towards better practice. ViH Online: Mar 10 2010 2:53AM DOI: 10.1111/j.1524-4733.2010.00700-x

10. Keeney RL, Raifa H. 1993. The multiplicative utility function. In Decisions with Multiple Objectives: Preferences and Value Tradeoffs. Cambridge University Press: Cambridge; 288-292.

11. Keeney RL, Raiffa H. 1976. Decision Making with Multiple Objectives. Wiley: New York.

12. Tversky A, Kaheman D. 1992. Advances in prospect theory: cumulative representation of uncertainty. Journal of Risk and Uncertainty 5: 297-323.

13. Fryback DG, Lawrence WF, Dollars may not buy as man QALYs as we think: A problem with defining quality of life adjustments, Med Decis Making 1997;17;276.

14. Manuel DG, Schultz SE, Kopec JA. Measuring the health burden of chronic disease and injury using health adjusted life expectancy and the Health Utilities Index. J Epidemiol Community Health 2002;56:843-850.

15. National Centre for Social Research and University College London. Department of Epidemiology and Public Health, Health Survey for England, 2006 [computer file]. 3rd Edition. Colchester, Essex: UK Data Archive [distributor], April 2010. SN: 5809.

16. National Centre for Social Research and University College London. Department of Epidemiology and Public Health, Health Survey for England, 2005 [computer file]. 2nd Edition. Colchester, Essex: UK Data Archive [distributor], April 2010. SN: 5675.
17. National Centre for Social Research and University College London. Department of Epidemiology and Public Health, Health Survey for England, 2004 [computer file]. 2nd Edition. Colchester, Essex: UK Data Archive [distributor], April 2010. SN: 5439.
18. National Centre for Social Research and University College London. Department of Epidemiology and Public Health, Health Survey for England, 2003 [computer file]. 2nd Edition. Colchester, Essex: UK Data Archive [distributor], April 2010. SN: 5098.
19. Dolan P, Gudex C, Kind P, Williams A. The time trade-off method: results from a general population study. Health Econ 1996;5:141-54.
20. Walters SJ, Brazier JE. Comparison of the minimally important difference for two health state utility measures: EQ-5D SF-6D. Qual Life Res 2005;14:1423-32.

Table 1 : Results from the OLS combination models

	Coefficient	Robust Std. Err.	P value	Coefficient	Robust Std. Err.	P value
Baseline:	Perfect Health			None of Health Conditions		
Minimum decrement	0.5136428	1.13	0.651	0.0439155	0.4978	0.93
Maximum decrement	0.5284501	1.3815	0.703	0.1545328	0.7076	0.828
CrossProduct	1.789911	1.7784	0.317	1.143514	0.8307	0.172
Constant	-0.6427511	1.6315	0.695	-0.1007821	0.7165	0.888
R Sq	0.5747			0.5803		

Minimum decrement: $\min(U_{nA} - U_A, U_{nB} - U_B)$
Maximum decrement: $\max(U_{nA} - U_A, U_{nB} - U_B))$
CrossProduct: $\min(U_{nA}, U_{nB}))*(U_A/U_{nA})*(U_B/U_{nB})$

Table 2: Comparing the accuracy of HSUVs generated using the five methods

	ACTUAL	Additive	Multiplicative	Minimum	ADE	OLS
Using a baseline of perfect health						
Mean EQ-5D score	0.5682	0.4288	0.5092	0.6667	0.6142	0.5669
Min EQ-5D score	0.3596	0.2321	0.3795	0.5860	0.5018	0.4367
Max EQ-5D	0.9165	0.7119	0.7284	0.7907	0.7777	0.8121
Range	0.5570	0.4797	0.3489	0.2047	0.2759	0.3754
Mean error		0.1384	0.0580	-0.0995	-0.0470	0.0003
Maximum error		0.3320	0.2129	0.2715	0.2206	0.1720
MAE		0.1411	0.0707	0.1037	0.0620	0.0471
RMSE		0.1529	0.0839	0.1214	0.0799	0.0603
Proportion within |0.01|		0%	7%	4%	15%	11%
Proportion within |0.05|		7%	39%	20%	46%	64%
Proportion within MID |0.074|		15%	58%	33%	72%	81%
Using age-adjusted baseline from individuals with none of health conditions						
Mean EQ-5D score	0.5682	0.4890	0.5418	0.6667	0.6367	0.5671
Min EQ-5D score	0.3596	0.2918	0.4040	0.5860	0.5266	0.4266
Max EQ-5D	0.9165	0.7532	0.7598	0.7907	0.7860	0.7955
Range	0.5570	0.4614	0.3558	0.2047	0.2595	0.3689
Mean error		0.0781	0.0254	-0.0995	-0.0695	0.0001
Maximum error		0.2792	0.1800	0.2715	0.2415	0.1732
MAE		0.0872	0.0516	0.1037	0.0781	0.0466
RMSE		0.1012	0.0651	0.1214	0.0950	0.0598
Proportion within |0.01|		7%	12%	4%	5%	13%
Proportion within |0.05|		26%	56%	20%	35%	63%
Proportion within MID |0.074|		40%	72%	33%	52%	80%

Table 3: Errors in estimated HSUVs sub-grouped by actual EQ-5D score

EQ-5D subgroup	Additive	Multiplicative	Minimum	ADE	OLS
ME: Using a baseline of perfect health					
Full set (n=97)	0.1384	0.0580	-0.0995	-0.0470	0.0003
1 to < 0.624 (n=23)	0.1730	0.1153	-0.0411	0.0007	0.0515
0.624 to < 0.562 (n=24)	0.1481	0.0739	-0.0723	-0.0236	0.0176
0.562 to < 0.514 (n=24)	0.1213	0.0403	-0.1069	-0.0550	-0.0151
0.514 to < 0.35 (n=26)	0.1145	0.0090	-0.1694	-0.1035	-0.0469
ME: Using age-adjusted baseline from individuals with none of health conditions					
Full set (n=97)	0.0781	0.0254	-0.0995	-0.0695	0.0001
1 to < 0.624 (n=23)	0.1119	0.0779	-0.0411	-0.0201	0.0530
0.624 to < 0.562 (n=24)	0.0871	0.0403	-0.0723	-0.0459	0.0141
0.562 to < 0.514 (n=24)	0.0609	0.0082	-0.1069	-0.0776	-0.0174
0.514 to < 0.35 (n=26)	0.0559	-0.0191	-0.1694	-0.1275	-0.0436
MAE: Using a baseline of perfect health					
Full set (n=97)	0.1411	0.0707	0.1037	0.0620	0.0471
1 to < 0.624 (n=23)	0.1730	0.1153	0.0579	0.0398	0.0596
0.624 to < 0.562 (n=24)	0.1481	0.0739	0.0730	0.0454	0.0341
0.562 to < 0.514 (n=24)	0.1234	0.0490	0.1069	0.0550	0.0375
0.514 to < 0.35 (n=26)	0.1226	0.0482	0.1694	0.1035	0.0570
MAE: Using age-adjusted baseline from individuals with none of health conditions					
Full set (n=97)	0.0872	0.0516	0.1037	0.0781	0.0466
1 to < 0.624 (n=23)	0.1123	0.0803	0.0579	0.0472	0.0604
0.624 to < 0.562 (n=24)	0.0871	0.0423	0.0730	0.0545	0.0320
0.562 to < 0.514 (n=24)	0.0739	0.0397	0.1069	0.0776	0.0392
0.514 to < 0.35 (n=26)	0.0775	0.0457	0.1694	0.1275	0.0548

Figure 1: Distribution of mean EQ-5D scores for subgroups with two comorbid health conditions

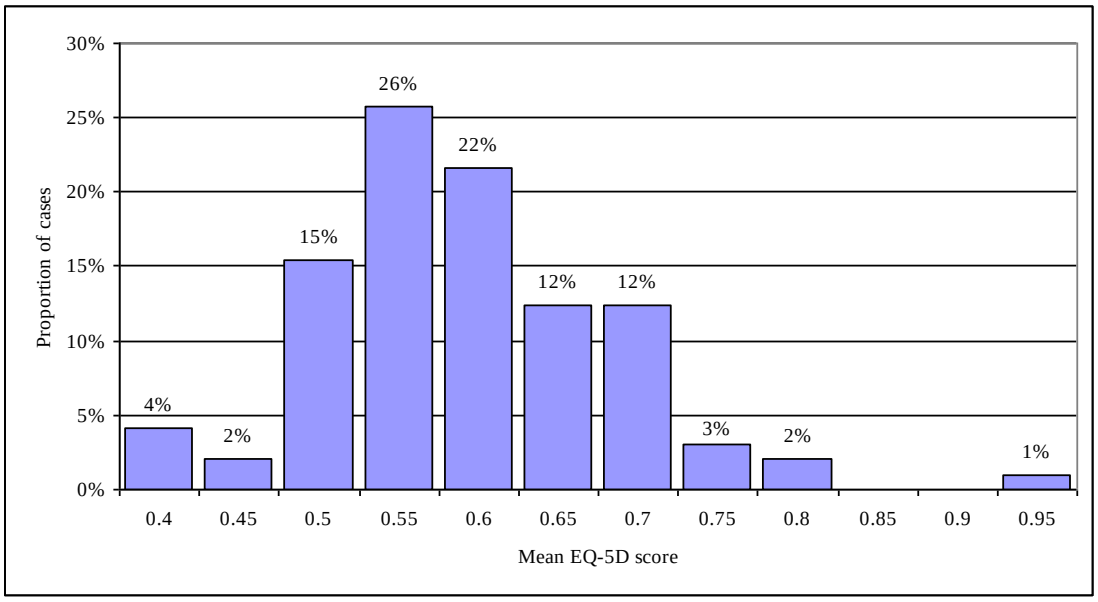

Figure 2a: Actual and estimated mean EQ-5D scores using the additive method

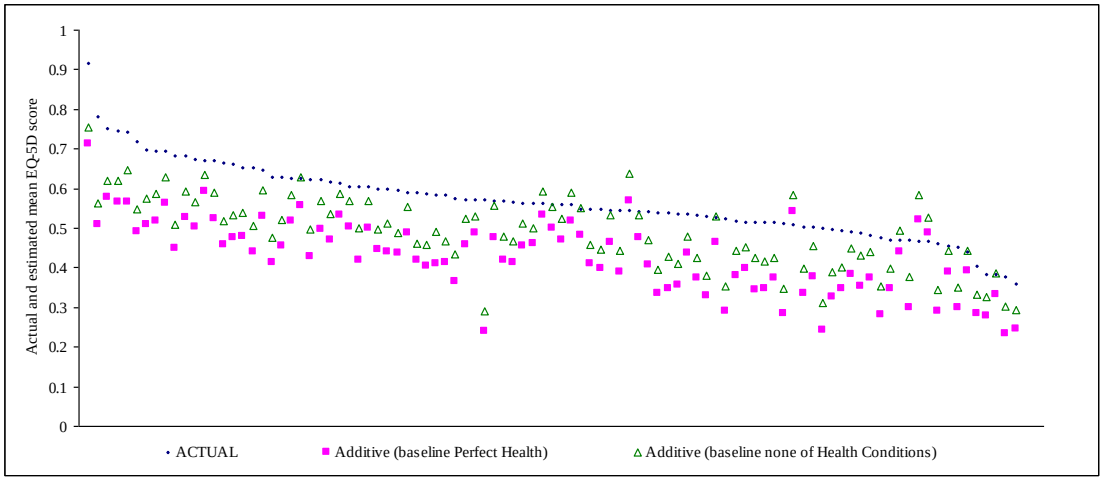

Figure 2b: Actual and estimated mean EQ-5D scores using the multiplicative method

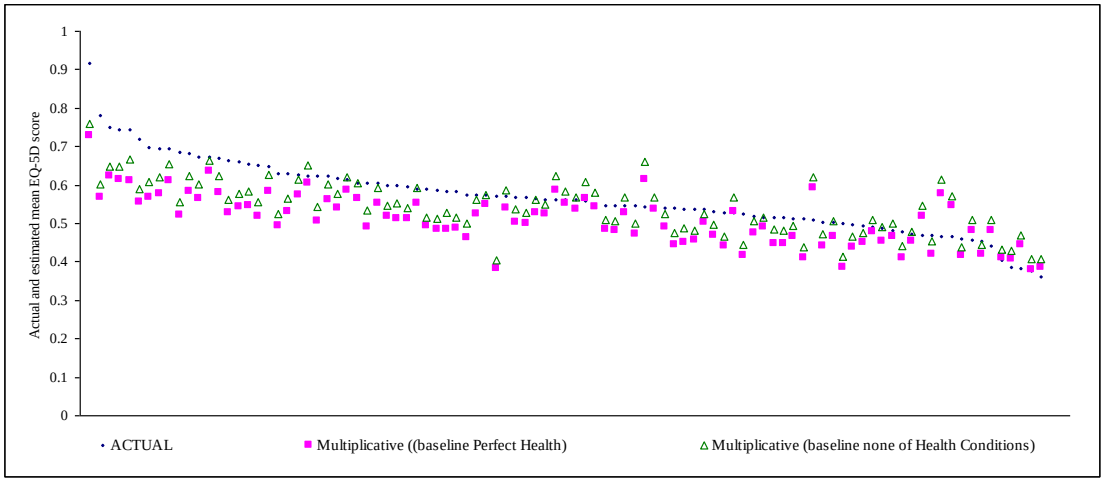

Figure 2c: Actual and estimated mean EQ-5D scores using the minimum method

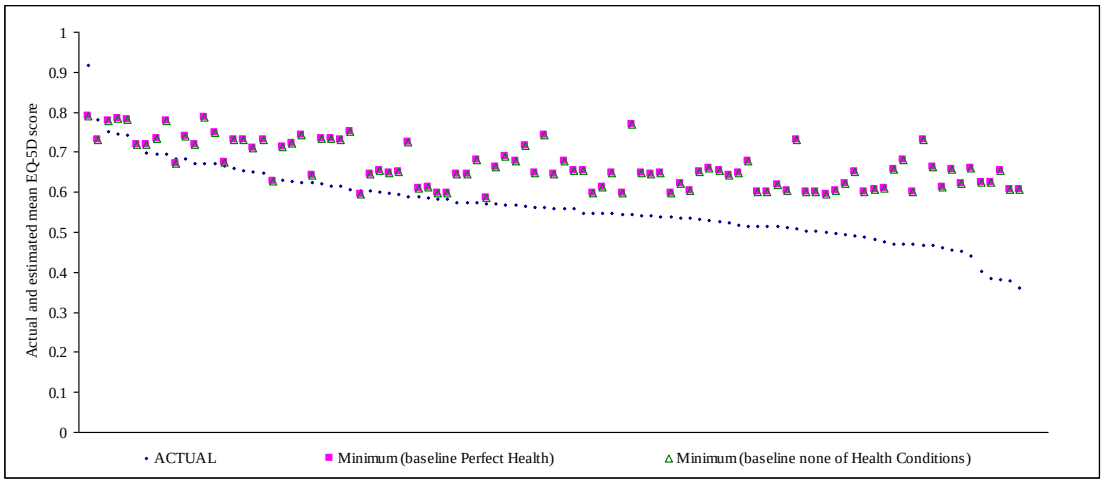

Figure 2d: Actual and estimated mean EQ-5D scores using the ADE method

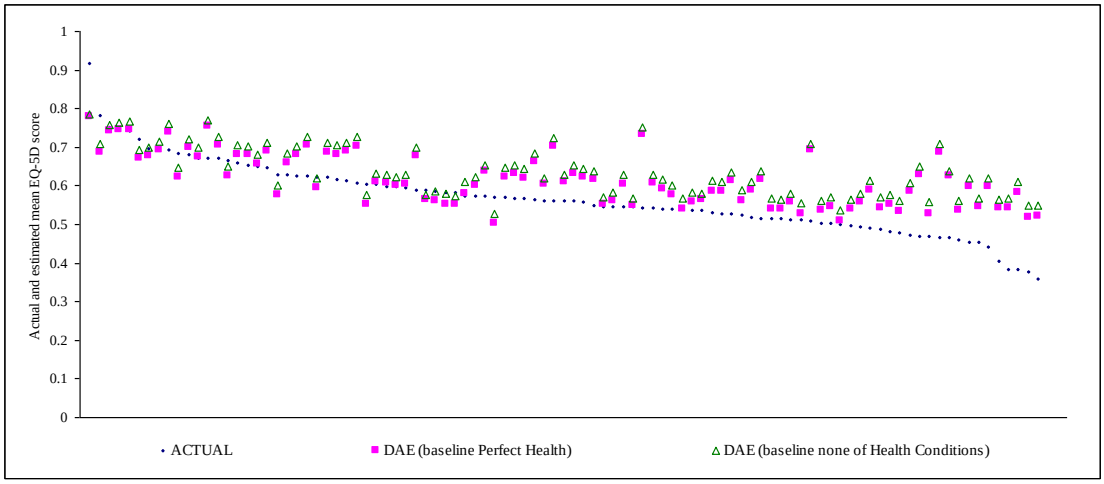

Figure 2e: Actual and predicted mean EQ-5D scores using the OLS regression model