HEDS Discussion Paper

No. 12.08

Health state valuation and mode of administration: Head to head comparison of online and CAPI

Brendan Mulhern (MRes)[1], Louise Longworth (PhD)[2], John Brazier (PhD)[1], Donna Rowen (PhD)[1], Nick Bansback (PhD)[3], Nancy Devlin (PhD)[4], Aki Tsuchiya (PhD)[1,5]

1. Health Economics and Decision Science, School of Health and Related Research, University of Sheffield
2. Department of Economics, University of Sheffield

Health state valuation and mode of administration:
Head to head comparison of online and CAPI

Brendan Mulhern (MRes)[1], Louise Longworth (PhD)[2], John Brazier (PhD)[1], Donna Rowen (PhD)[1], Nick Bansback (PhD)[3], Nancy Devlin (PhD)[4], Aki Tsuchiya (PhD)[1,5]

1. Health Economics and Decision Science, School of Health and Related Research, University of Sheffield, Regent Court, Sheffield, S1 4DA, UK
2. Health Economics Research Group, Brunel University, UB8 3PH
3. School of Population and Public Health, University of British Columbia, Vancouver, BC, V6Z 1Y6
4. Office of Health Economics, Southside 7th floor,105 Victoria Street, London SW1E 6QT
5. Department of Economics, University of Sheffield, Sheffield, S1 4DT

Corresponding author:
Brendan Mulhern, School of Health and Related Research, University of Sheffield, Regent Court, 30 Regent St, Sheffield, S1 4DA, United Kingdom
e-mail: b.mulhern@sheffield.ac.uk
Tel no: +44 (0)114 222 0794
Fax no: +44 (0)114 272 4095
Running head: Health state valuation: Online and CAPI comparison

Funding:
This study is funded by the MRC Methodology programme (ref.G0901500; Project title: 'Preparatory study for the Revaluation of the EQ-5D Tariff' (PRET))

Key words: utility assessment, online, CAPI, health economics methods, health related quality of life
Abstract

Abstract

Objectives: Health state valuation exercises can be conducted online but the quality of data generated is unclear. We report a study investigating whether responses to binary choice health state valuation questions differ by administration mode (online vs. face-to-face).

Methods: Identical surveys including demographic, self reported health and seven types of binary choice valuation questions were administered in online and Computer Assisted Personal Interview (CAPI) settings. Samples were recruited following procedures employed in typical online or CAPI studies. Analysis included descriptive comparisons of the distribution of responses across the binary options, and probit regression to explain the propensity to choose one option across modes of administration, controlling for background characteristics.

Results: Overall, 422 (221 online; 201 CAPI) respondents completed a survey. There were no overall age or gender differences. Online respondents were educated to a higher level than the CAPI sample and general population, and employment status differed. CAPI respondents reported significantly better general health, and health/life satisfaction. CAPI took significantly longer to complete. There was no effect of mode of administration on responses to the valuation questions, and this was replicated when demographic differences were controlled.

Conclusions: The findings suggest that both modes may be equally valid for health state valuation studies using binary choice methods (e.g. Discrete Choice Experiments). There are some differences between the observable characteristics of the samples, and the groups may differ further in terms of unobservable characteristics. When designing health state valuation studies, the advantages and disadvantages of both approaches must be considered.

Introduction

To conduct health state valuation studies, a range of administration modes can be used. These include face-to-face interviews using paper and pencil methods, face-to-face interviews using computer assisted personal interviews (CAPI – where surveys are displayed via a computer interface while an interviewer is present), and online studies. To date, face-to-face interviews with pencil and paper have been the most widely used mode for collecting health state valuation data, and were used to derive preference weights for EQ-5D using the Time Trade-Off (TTO) preference elicitation technique [1] and SF-6D using the Standard Gamble (SG) approach [2,3]. TTO studies employing CAPI methods have also been used to derive EQ-5D value sets [4,5]. An online version of the iterative TTO process has been investigated, and some concerns have been found with the approach [6,7]. Preferences have also been elicited in a face-to-face setting using discrete choice experiments (DCE) [8], and recently DCE preference elicitation studies have been conducted using online methods [6].

Each mode has advantages and disadvantages that may impact on the data generated and therefore need to be considered in the design of health state valuation studies. Face-to-face interviews may provide high quality data with good completion rates and reliability [9], but are expensive and time consuming to conduct. The use of CAPI means some of the advantages of online surveys can be exploited in face-to-face environments. For example, CAPI interviews allow for complex routing of questions, question order randomisation, recording of the time taken, and the minimisation of errors associated with data entry (which is completed automatically). Online valuation methods have the advantages related to CAPI but in addition are cheaper to conduct, allow large samples to be achieved in a short space of time, have a flexible sampling frame, and enable a range of background characteristics of non-respondents to be obtained.

Studies have compared the online and CAPI administration of health state valuation exercises. Comparisons of the online and face-to-face administration of iterative Person Trade Off (PTO) tasks have found broadly similar results across modes [10, 11]. However Robinson and colleagues [10] found that a greater number of online respondents gave ‘equivalence’ responses, which provided the quickest way to finish each PTO task. Norman and colleagues [7] have compared the online and face-to-face (but not CAPI based) administration of an iterative TTO task, and found that the responses differed by administration mode, with those completing the online survey displaying more variation in response. When the results were modelled to generate a tariff for EQ-5D, 100 of the 243 health state values were higher in the online group to the order of at least 0.1. Therefore, iterative health state valuation tasks administered online may generate different results from CAPI, but it is not clear whether the difference comes from the mode of administration or an interaction between the iterative task and the mode of administration.

It has also not been established whether responses to binary choice health state valuation questions are comparable across administration modes. Binary choice questions such as DCE_{TTO}, where duration is included as a dimension alongside the health state classification system can be used to value health states at the aggregate level [6]. Binary choice questions such as DCE with no duration can be used to value health states in "hybrids" with TTO [12,13]. Since conventional health state valuation questions such as SG or TTO based on iteration to identify indifference are effectively made up of a series of binary choice questions, a set of independent 'snapshot' binary choice questions can be used to value health states at the aggregate level [14]. Testing the online vs. CAPI administration of binary choice questions to investigate the level of equivalence between the modes of administration has direct implications for the better understanding of these methods, and also the design of future health state valuation studies using binary choice techniques.

Collecting data online raises concerns about the representativeness of the sample, and comparability with data collected using other methods. Samples recruited online may be biased in terms of unobserved characteristics [15]. There are also concerns about the motive of participation, the level of non-response and attrition [16], the reliability and validity of the data generated [17-19], and the level of engagement of respondents. However, other studies have found comparability across samples [20,21], and also comparability in terms of the reliability and validity of the data generated [22].

This study aims to compare responses to binary choice versions of health state valuation questions across online and CAPI modes of administration. This is done by administering identical surveys where the only difference is the mode of administration used to collect the data. We hypothesise that responses to the binary choice valuation questions do not differ across modes. To investigate issues around the representativeness of sample recruited to face-to-face and online studies, we also compare the demographic characteristics and self reported health status of the samples, and the general population.

Methods

Survey

The survey used in this study to test for differences between the online and CAPI modes of administration contained identical demographic and binary choice questions in the same order. The binary choice questions were used to investigate methodological issues related to health state valuation, and a full description of the issues investigated and the binary choice questions used is provided elsewhere [23]. Seven different "types" of binary choice questions were included in the survey, and the basic format of the questions is described in Figure 1. The questions include a number of experimental attributes that can be varied across hypothetical health states. These include health state experienced (H), time spent in the health state (T), lead time spent in full health prior to the health state (L), perspective of the person experiencing the health state (P) and level of

satisfaction with health whilst experiencing the health state (S). Not all of the attributes vary across each question (see Table 1 for the attribute combinations used across each question type, and Tsuchiya & Mulhern [23] for more details). Overall, 12 questions were administered across three question modules (the time taken to complete each module was recorded). Module 1 included five type I questions, module 2 included one of question types II-VI (so five questions in total), and module 3 included two DCE_{TTO} questions (type VII). The health states used were based on the EQ-5D-5L [24]. EQ-5D-5L describes health status across five dimensions (mobility, self care, usual activities, pain/discomfort and anxiety/depression) each with five response levels (no, slight, moderate, severe and extreme problems).

Each survey began by providing study information, and this was followed by a compulsory informed consent page. Respondents were then asked a series of demographics and completed self reported health status (on a five point scale from "excellent" to "poor"), health and life satisfaction questions (on a ten point scale from "completely satisfied" to "completely dissatisfied") and EQ-5D-5L before completing the twelve binary choice valuation questions.

Recruitment and the sample

To achieve a comparison of the two modes of administration as they would happen in the real world, the CAPI and online samples were recruited separately following procedures employed in typical surveys.

For the online survey, respondents were sourced from an existing internet panel, and were selected following set quotas based on the UK general population across five age groups (18-24; 25-34; 35-44; 45-54; 55-64) and gender. Invitations were sent out by e-mail. Potential respondents were screened out prior to starting the experimental questions if the relevant quota for age and gender was already complete, or after the completion if they completed the survey in less than the minimum imposed time limit of five minutes. The online survey described in this study was one of 15 different online surveys that aimed for an overall achieved sample of 3000 (approx 200 respondents per version). The questions included in the version described here were identical to those used in the CAPI survey, and therefore a subset of respondents from the overall sample is considered in this study.

For the CAPI interviews, recruitment followed set quotas for age and gender based on the UK general population and scaled down for an achieved sample of 200. This attempted to ensure overall comparability of the sample characteristics across the administration modes. Participants were recruited by knocking on one in every ten doors in randomly selected postcodes in five UK areas. The survey was presented to respondents on a laptop, and the interviewer read out all of the questions and text on the screen, and recorded the response given by the respondent. This was done in a one-to-one setting. The same minimum completion time of 5 minutes was imposed, and participants were able to stop the survey at any time.

Analysis

Background characteristics, self reported health and time taken to complete the survey were compared across the two samples using chi squared and ANOVA analyses. The background characteristics of the overall, CAPI and online samples were also compared with the general population of England and Wales using statistics extracted from the 2001 UK census [25] for 18-64 year olds. Comparisons of the proportion of respondents who choose scenario B, by sample, and by the seven binary choice question types were also carried out, with statistical significance indicated by p values < 0.05.

Probit regressions were used to explore the determinants of the propensity to choose scenario B for each question:

$$\Pr(B = 1) = \Phi\left(\boldsymbol{\beta}_1\mathbf{D} + \boldsymbol{\beta}_2\mathbf{S} + \boldsymbol{\beta}_3\mathbf{X}\right)$$

where Pr represents probability, the β_i's are parameters to estimate, D represent the socio-demographic characteristics of respondents, S represents health satisfaction of the respondent, X represents the properties of the health state using health state (H), duration (T), lead time in full health (L), person perspective (P), and satisfaction level (S) (see figure 1 and table 1), and the function $\Phi(.)$ is the distribution function of the standard normal distribution [26,27]. Marginal effects are reported as they can be interpreted as percentages. E.g. a marginal effect of -0.2 for male indicates that being male reduces the probability of choosing B by 20%. For the regressions, statistical significance levels of both <0.05 and <0.1 are reported.

Results

Respondent characteristics and self reported health

In total, 422 respondents completed either the online survey or the CAPI version. For the online survey 2326 panel members were invited to take part and 487 potential respondents (20.1%) accessed the survey. Of these 266 (11% of those invited; 54% of those accessing the survey) were screened out (as they were of an age and gender quota that was already complete), left the survey or completed the survey in less than 5 minutes so were defined as non completers, and 221 (9.5% of those invited (46% of those accessing) fully completed the survey in five minutes or more. There were no significant age or gender differences between the responder and non responder samples. The CAPI version was completed by 201 respondents. The number of respondents invited to take part is not available, and therefore the response rate cannot be calculated. No CAPI respondents were excluded for completing the survey too quickly, and no respondents asked to stop the survey once they had begun answering the questions.

There were no significant differences between the online and CAPI groups by age and gender, but a number of demographic variables significantly differed between the samples (Table 2). These

included employment status (with more retired people and homemakers in the CAPI sample, but more students in the online sample), marital status (with more CAPI respondents being married and more online respondents being single) and education level (with online respondents being educated to a higher level). In comparison to the general population, the overall, online and CAPI samples all differ in terms of employment status, with more employed people in the general population (all $p < 0.001$). The online sample is more similar to the general population in terms of marital status than the CAPI sample (with more of the CAPI sample being married or with a partner). However, the overall education level of the CAPI sample is more similar to the general population.

The CAPI sample took significantly longer to complete the overall survey, and also Module 1 (five type I binary choice questions) and module 2 (one of types II- VI binary choice questions). There were no differences between the samples for the time taken to complete module 3 (two DCE_{TTO} questions). Across all three modules the standard deviation of the time taken is longer for the online than CAPI sample.

Responses to the self report general health and health and life satisfaction questions are displayed in Figure 2. The CAPI sample are significantly more likely to report better health ($p = 0.002$), higher levels of health satisfaction ($p < 0.001$), and higher levels of life satisfaction ($p < 0.001$). The mean EQ-5D-5L index score for the online sample (mapped from EQ-5D-3L using the algorithm produced by van Hout et al [28]) was 0.776 (0.25) and for the CAPI sample was 0.874 (0.20). This difference was significant ($F(1,409) = 18.66$, $p < 0.001$). EQ-5D-5L dimensions responses also differ significantly by mode of administration with the exception of mobility, with the CAPI group reporting less problems.

Binary choice valuation questions

The proportion of the sample choosing scenario B (i.e. choosing to live for a shorter duration in full health or choosing immediate death) did not significantly differ by administration mode for any of the seven binary choice question types. This was irrespective of the experimental attributes varied in the scenario (Table 3).

Probit regressions for each question reveal that a range of demographic and experimental attribute variables significantly predict the likelihood of choosing scenario B for a number of the binary choice questions, but mode of administration does not significantly predict response across any of the question types (Table 4). For Type I questions, response is significantly predicted by the health state and duration used in the question, where the more severe the health state or the larger the duration, the more likely scenario B is selected. These results cannot be tested across the other question Types, as Types II-VI include only one health state and associated duration. For Type II questions, females are 4% more likely and those with higher levels of life satisfaction are 1% more likely to choose to live in full health. For Type IV, females are 8% more likely to choose scenario B, and for

Type V, males are 10% more likely and respondents who are retired are 19% more likely to choose scenario B. Response to Type VII questions is predicted by education level and life satisfaction, but these results are difficult to interpret due to the nature of the Type VII questions which presents two full EQ-5D-5L health states. Response to question Types III and VI is not predicted by any of the variables.

Discussion

When health state valuation techniques such as TTO and SG were developed, face-to-face interviews were seen as the best way to administer the exercises, and this is the mode used to derive preferences for generic preference based measures of health such as EQ-5D [1] and SF-6D [2,3], and also condition specific instruments [29-31]. In recent years there have been advances in communication technology and interest in the use of online health state valuation techniques is increasing. In parallel to this, health state valuation methods that are amenable to online administration, based on binary choice questions, have been developed [6,14,32]. However, there are issues regarding the quality of data generated using online surveys. This paper reports on a comparison between an identical set of binary choice questions designed to test issues related to health state valuation conducted in online and face-to-face environments. The results support the hypothesis that there is no difference between the responses to the valuation tasks across the administration modes. We also investigated the sample characteristics, and found some differences, between the groups. However the finding of equivalence across the modes remained robust after controlling for differences in the sample characteristics.

The responses to the main binary choice valuation questions were not statistically significantly different across the modes of administration, and this finding was not influenced by the severity or duration of the state used in the binary choice question. This is in comparison to a study comparing an iterative valuation technique (TTO) that found differences between online and face to face responses, and concluded that this was due to the iterative nature of the process [7]. This is because respondents that intend to complete the questions quickly may accept the first trade off offered to avoid going through the process to reach indifference. Our study did not test an iterative process, but rather binary choice health state valuation tasks, which are amenable to online and CAPI administration. The outcome has been that where a design that is suited to online and CAPI administration is used, the hypothesis that the mode of administration does not impact on the results is supported, as comparable results are generated. This finding is valid for samples recruited following the standard procedures for CAPI (i.e. knocking on doors in selected postcode areas to produce a representative sample) and online (i.e. using participant panels) who were found to have similarities with the general population (the group targeted in most preference elicitation studies conducted the world [1-6]). This demonstrates the potential applicability of our results in the design of valuation studies using binary methods. However it is unclear how these findings relate to other preference elicitation tasks. It may be possible to extend our findings to other valuation methods, and

further work should consider the stability of a range of both iterative and binary choice preference elicitation techniques across different administration modes.

We have also assessed the time taken to complete the survey. If an online respondent completes the survey too quickly or too slowly, this may suggest that they are not fully engaged. The results demonstrate that the CAPI sample took significantly longer to complete the overall survey, and two of the experimental modules. This may be because an interviewer is present and reads out the questions, and it is unlikely for the respondent to complete the survey without some minimal level of engagement. The shorter completion time data indicate that it is possible that at least some respondents in the online sample completed the survey without fully paying attention or engaging in the task, and future research may wish to investigate respondent engagement in the online environment in more detail, for example by recording the time taken to complete each task, or developing innovative methods for presenting the tasks.

There have been concerns about the representativeness of online samples and how this might impact the comparability of results across samples [15,20,21]. The two samples in our study were recruited against age and gender quotas and therefore do not differ in terms of these characteristics. However the two samples differ significantly in some observable characteristics and this raises the issue of representativeness with respect to the UK general population. Compared to previous census data [25], the CAPI sample are more representative in terms of educational attainment. The online sample over-represents people educated to at least degree level, and this has also been found in other studies comparing online research groups to the general population [20]. Those who are educated to a higher level may be more computer literate and this possibly explains their over-representation in the sample. In terms of health, it is possible that the online sample is genuinely less healthy than the CAPI sample. However, it has also been established that individuals may answer face-to-face surveys in a socially desirable way, particularly when answering questions about sensitive issues such as mental health [33]. This may vary according to whether responses were public or anonymous [34]. In the CAPI sample there may be a discrepancy between actual health and reported health status because of the presence of the interviewer, which may mean that, from the respondent's perspective, responses are not completely anonymised. This did not, however, impact on responses to the health state valuation questions.

We were not able to assess how mode of administration impacts the responses of those aged over 65, as this group were not included in the sampling frame for the study. This potentially limits the applicability of our findings, as the ONS predicts that approximately 59% of adults aged 65 or older use the internet every day or almost everyday (in comparison to approximately 80% of those aged 18-54, and 75% of those aged 55-64) [35]. Robinson et al [10] included members of the population aged over 65 in their comparison of PTO tasks across different modes of administration, and found that response to the internet arm of the study was lower in this group. However they did not investigate

differences in responses to the PTO task across different age groups. Further comparisons of valuation tasks across different modes of administration should investigate responses amongst those aged over 65. This will establish the level of equivalence of health state valuation exercises across different modes of administration for the overall adult population.

It is in theory possible to make two samples agree in terms of any observable characteristic. However, even with highly selective screening, the samples may differ in terms of further unobserved characteristics. The CAPI sample characteristics are influenced by who is at home when the interviewer visits, who agrees to take part, and who completes the interview. The online sample using an internet panel is impacted by who has access to the internet, who is a member of the online panel, who in the panel agrees to take part, and who of those agreeing to take part completes the survey. It is not clear how the different selection mechanisms impact on unobservable sample characteristics, and therefore on responses to health state valuation questions. Not everyone is equally likely to join an online panel and complete a particular survey, or take part in a face-to-face interview. Typically characteristics of non-responders to interviews are not available, and one advantage of online surveys using existing internet panels is that certain characteristics of non responders may be accessible. This allows for further insight into issues around non response.

When designing health state valuation studies, the financial and time costs of the surveys are important and must be considered in light of the available preference elicitation methods and the modes available for administering the techniques. The cost of any survey has a fixed element and is not completely proportionate to sample size. However, even if generalised cost estimates cannot be given, it is generally the case that face-to-face interviews are substantially more costly per respondent and may take a much longer time to recruit sufficient numbers of participants. At the same time, any survey is only as good as the quality of the sample, and therefore the quality of the internet panel needs to be scrutinised: one way to do this may be to look at how panel members are recruited and what incentives are offered. Overall, when the survey design is amenable to online administration, the incremental cost effectiveness of conducting interview surveys must be examined.

This paper discusses the findings from a head to head comparison of online and CAPI administrations of binary choice health state valuation questions. The two administrations have different advantages and disadvantages, and the two samples significantly differ across selected background characteristics, but the similarities with the general population indicate that the standard sampling frames used for face to face and online research studies are valid. However, responses to the main experimental binary choice questions were not significantly different across the modes, and mode of administration was not a significant factor explaining the responses. Therefore, both modes produce comparable data, and both can be used to administer health state valuation surveys including binary choice valuation questions such as DCE_{TTO} [6]. The advantages and disadvantages of both modes must be considered when designing health state valuation studies.

Acknowledgements

The wider PRET project team consists of Nick Bansback, John Brazier, Ken Buckingham, John Cairns, Nancy Devlin, Paul Dolan, Arne Risa Hole, Louise Longworth, Brendan Mulhern, Donna Rowen, and Aki Tsuchiya (Principal Investigator). We are grateful to all respondents who took part, and the market research companies for carrying out the data collection. The study has been approved by the ethics committee at the University of Sheffield School of Health and Related Research. The usual disclaimer applies.

References

1 Dolan P. Modelling valuations for EuroQol health states. Med Care. 1997;35(11):1095-108.

2 Brazier J, Roberts J, Deverill M. The estimation of a preference-based measure of health from the SF 36. J Health Econ. 2002;21(2):271-92.

3 Brazier JE, Roberts J. Estimating a preference-based index from the SF-12. Med Care. 2004;42(9):851-59.

4 Wittrup-Jensen KU, Lauridsen J, Gudex C, Pedersen KM. Generation of a Danish TTO value set for EQ-5D health states. Scandinavian Journal of Public Health. 2009;37:459-66.

5 Viney R Norman R, King MT, et al. Time Trade Off EQ-5D weights for Australia. Value Health. 2011;14:928-936.

6 Bansback N, Brazier J, Tsuchiya A, Anis A. Using a discrete choice experiment to estimate societal health state utility values. J Health Econ. 2012; 31(1):306-318.

7 Norman R, King MT, Clarke D, et al. Does mode of administration matter? Comparison of online and face-to-face administration of a time trade-off task. Qual. Life Res. 2010;19(4):499-508.

8 Brazier J, Rowen D, Yang Y, Tsuchiya A. Comparison of health state utility values derived using time trade-off, rank and discrete choice data anchored on the full health-dead scale. Eur J Health Econ. 2011;DOI10.1007/s10198-011-0352-9.

9 Dolan P, Gudex C, Kind P, Williams A. Valuing health states: A comparison of methods. J Health Econ. 1996;15:209-31.

10 Robinson A, Covey J, Jones-Lee M, Loomes G. Comparing the results of face-to-face and web based PTO exercises. Health Economics Study Group; 2008; East Anglia, UK.

11 Damschroder LJ, Baron J, Hershey JC, et al. The validity of person tradeoff measurements: Randomized trial of computer elicitation versus face-to-face interview. Med Decis. Mak. 2004;24:170-80.

12 Rowen D, Brazier J, van Hout B. A comparison of methods for converting DCE values onto the full health-dead QALY scale. HEDS Discussion Paper. 2011;11/15.

13 van Hout B, Oppe M. Combining DCE and TTO into a single value function. Value Health. 2011;14(7):A238.

14 Bosch JL, Kammitt JK, Weinstein MC, Hunink MGM. Estimating general-population utilities using one binary-gamble question per respondent, Med. Decis. Mak 1998;18:381-390.

15 Schmidt WC. World Wide Web survey research: benefits, potential problems, and solutions. Behavior Research Methods, Instruments and Computers, 1997;29:274-79.

16 Eysenbach G. The law of attrition. Journal of Medical Internet Research. 2005;7(1):e11.

17 Bowling A. (2005). Mode of questionnaire administration can have serious effects on data quality. Journal of Public Health, 27(3): 281-291.

15 Rieps U. The web experiment method: advantages, disadvantages, and solutions. In: Birnbaum MH, editor. Psychological experiments on the internet. San Diego, CA: Academic Press, 2000; p. 89-117.

19 Strickland OL, Moloney MF, Dietrich AS, et al. Measurement issues related to data collection on the World Wide Web. Adv. Nurs. Sci. 2003;26:246-56.
20 Liu H, Cella D, Gershon R, et al. Representativeness of the Patient-Reported Outcomes Measurement Information System Internet Panel. J. Clin. Epidemiol. 2010;63:1169-178.
21 Schillewaert N, Meulemeester P. Comparing response distributions of offline and online data collection methods. International Journal of Market Research. 2005;47:163-78.
22 Raat H, Mangunkusumo RT, Landgraf JM,et al. Feasibility, reliability, and validity of adolescent health status measurement by the Child Health Questionnaire Child Form (CHQ-CF): Internet administration compared with the standard paper version. Qual. Life Res. 2007;16:675-85.
23 Tsuchiya A, Mulhern B. Preparation for the Re-valuation of the EQ-5D Tariff (PRET) project: Overview of methods for project stages 1-3. HEDS Discussion Paper. 2011;11/16.
24 Herdman M, Gudex C, Lloyd A, et al. (2011). Development and preliminary testing of the new five-level version of EQ-5D (EQ-5D-5L). Qual. Life Res. 2011;20(10):1727-36
25 Office of National Statistics. Census 2001: General report for England and Wales. UK: ONS; 2005.
26 Alexandre PK, French MT 'Further Evidence on the Labor Market Effects of Addiction: Chronic Drug Use and Employment in Metropolitan Miami', Contemporary Economic Policy. 2004;22:382-93.
27 Greene WH. Econometric Analysis. New Jersey: Prentice Hall; 2000.
28 van Hout B, Janssen MF, Feng YS, Kohlmann T, Busschbach J, Golicki D, Lloyd A, Scalone L, Kind P, Pickard AS. Interim scoring for the EQ-5D-5L: Mapping the EQ-5D-5L to EQ-5D-3L value sets. Value Health; doi:10.1016/j.jval.2012.02.008.
29 Rowen D, Mulhern B, Banerjee S, et al. Estimating preference based single index measures for dementia using DEMQOL and DEMQOL-Proxy. Value Health. 2012; 15(2):346-356.
30 Yang Y, Brazier J, Tsuchiya A, Coyne K. (2008) Estimating a preference-based index from the Overactive Bladder questionnaire. Value Health. 2008;12:59–66.
31 Yang Y, Brazier JE, Tsuchiya A, Young T. (2011) Estimating a preference based index for a 5-dimensional health state classification for asthma derived from the asthma quality of life questionnaire. Med Decis Mak. 2011;31:281-91.
32 Ratcliffe J, Couzner L, Flynn T, et al. Valuing Child Health Utility 9D health states with a young adolescent sample: a feasibility study to compare Best-Worst Discrete Choice Experiment, Standard Gamble and Time Trade Off methods. Applied Health Economics and Health Policy. 2011;9(1):15-27.
33 Robling MR, Ingledew DK, Greene G, et al. Applying an extended theoretical framework for data collection mode to health services research. BMC Health Services Research. 2010;10:180.
34 Paulhus DL. Two-component models of socially desirable responding. J Pers. Soc. Psychol. 1984;46(3):598-609.
35 Office of National Statistics. Internet Access 2010: Households and Individuals. UK: ONS; 2010.

Table 1: The 12 experimental questions used for the survey

Type	Scenario A					Scenario B	
	H	T	L	P	S	H	T
I	Slight problems walking about	10 years	n/m	You	n/m	Full health	9 years
I	Slight pain	10 weeks	n/m	You	n/m	Full health	8 weeks
I	Unable to walk about	10 years	n/m	You	n/m	Full health	8 years
I	Extreme pain	2 years	n/m	You	n/m	Full health	5 years
I	Extremely depressed	1 year	n/m	You	n/m	Full health	7 months
II	Extreme pain	10 years	n/m	Somebody else	n/m	Full health	6 years
III	Slight pain	10 weeks	10 weeks	You	n/m	Full health	19 weeks
IV	Extremely depressed	1 year	10 weeks	Somebody else like you	n/m	Full health	7 months
V	Unable to walk about	5 years	n/m	You	High	Full health	3 years
VI*	55555	10 years	10 years	You	n/m	Immediate death	n/a
VIIa*	24144	5 years	n/m	You	n/m	54514	1 year
VIIb*	25555	1 year	n/m	You	n/m	42424	1 year
VIIc*	53543	10 years	n/m	You	n/m	31354	10 years
VIId*	41234	1 year	n/m	You	n/m	14112	1 year

* EQ-5D-5L health state listed; n/m: not mentioned in scenario

Table 2: Sample characteristics

Characteristic	Overall	Online	CAPI	General population	P value (online, CAPI and GP)	P value (online vs. CAPI)	P value (overall vs. GP	P value (online vs. GP)	P value (CAPI vs. GP)
n	422	221 (52.37)	201 (47.63)	Matched					
Age									
Mean (SD)	41.49 (13.96)	41.56 (14.38)	41.41 (13.52)	42.23	n/a	*P* = 0.913	n/a	n/a	n/a
Range	18-65	18-65	18-65	18-64					
Age category (n,%)					*P* = 0.411	*P* = 0.233	*P* = 0.415	*P* = 0.154	*P* = 0.980
18-24	64 (15.2)	34 (15.4)	30 (15.0)	13.7					
25-34	97 (23.0)	51 (23.1)	46 (22.9)	23.2					
35-44	85 (20.1)	36 (16.29)	49 (24.4)	24.3					
45-54	86 (20.4)	46 (21.8)	40 (19.9)	21.6					
55-64	90 (21.3)	54 (24.4)	36 (17.9)	17.2					
Male (n,%)	201 (47.6)	102 (46.2)	99 (49.3)	47.9	*P* = 0.836	*P* = 0.524	*P* = 0.945	*P* = 0.703	*P* = 0.765
Employment (n,%)					*P* < 0.001	*P* = 0.009	*P* < 0.001	*P* < 0.001	*P* < 0.001
In employment	245 (58.1)	128 (57.9)	117 (58.2)	70.3					
Retired	41 (9.7)	17 (7.7)	24 (11.9)	4.7					
Homemaker	34 (8.1)	10 (4.5)	24 (11.9)	7.5					
Student	36 (8.5)	23 (10.4)	13 (6.5)	1.9					
Seeking work	16 (3.8)	10 (4.5)	6 (3.0)	n/a					
Unemployed	18 (4.3)	9 (4.1)	9 (4.5)	3.7					
Long term sick	25 (5.9)	18 (8.1)	4 (2.0)	5.8					
Other	7 (1.7)	6 (2.7)	1 (0.5)	3.4					
Marital status (n,%)					*P* = 0.047	*P* = 0.013	*P* = 0.297	*P* = 0.705	*P* = 0.044
Married/partner	236 (55.9)	111 (50.7)	125 (62.2)	52.6					
Single	184 (43.6)	108 (49.3)	76 (37.8)	47.4					
Education cont after minimum age (n,%)	292 (69.2)	174 (78.7)	118 (58.7)	n/a	n/a	*P* < 0.001	n/a	n/a	n/a
Educated to degree level (n,%)	136 (29.9)	90 (40.7)	46 (22.9)	21.6	*P* < 0.001	*P* = 0.032	*P* < 0.001	*P* < 0.001	*P* = 0.719
Time taken to complete (M (sd) minutes)									
Overall	9.88 (4.6)	8.64 (3.84)	11.26 (4.99)	n/a	n/a	*P* < 0.001	n/a	n/a	n/a
Module 1	1.27 (0.76)	1.07 (0.77)	1.49 (0.70)	n/a	n/a	*P* < 0.001	n/a	n/a	n/a
Module 2	1.92 (1.33)	1.80 (1.63)	2.06 (0.89)	n/a	n/a	*P* = 0.045	n/a	n/a	n/a
Module 3	1.28 (0.99)	1.20 (1.11)	1.36 (0.84)	n/a	n/a	*P* = 0.088	n/a	n/a	n/a

Nb: General population of England and Wales extracted for 18-65 years olds from the 2001 census (ONS, 2005)

Table 3: Proportion of respondents choosing Scenario B in different binary choice questions

Type	Online (%)	CAPI (%)	P value
I	67.0	68.2	0.79
I	54.8	58.7	0.41
I	81.9	81.6	0.94
I	98.2	98.5	0.80
I	91.9	91.5	0.91
II	92.8	94.0	0.60
III	71.0	75.6	0.29
IV	81.9	83.1	0.75
V	56.6	60.7	0.39
VI	65.6	64.6	0.84
VII	49.1	49.8	0.91
VII	77.8	76.6	0.82

Question types I-V, Scenario B represents living in full health for a shorter duration.
Question type VI, Scenario B represents immediate death.
Question type VII, Scenario B is a 5 level EQ-5D-5L health state with associated duration.

Table 4: Probit marginal effects coefficients for the likelihood of choosing scenario B

Variable	Type I	Type II	Type III	Type IV	Type V	Type VI	Type VIIa	Type VIIb
Health state	0.09*	-	-	-	-	-	-	-
V value	0.05*	-	-	-	-	-	-	-
Duration	0.03*	-	-	-	-	-	-	-
Administration mode	0.00	-0.00	0.05	0.00	0.03	-0.05	0.05	0.01
Gender	0.01	0.04**	-0.01	0.08*	-0.10*	0.04	-0.04	-0.01
Age	-0.01**	0.00	0.02	-0.02	-0.02	-0.02	0.01	0.02
Education level	0.03**	0.01	0.00	-0.05	-0.01	0.04	-0.09	-0.10**
Health status	-0.00	-0.01	-0.05	-0.04	-0.02	-0.03	-0.03	0.04
Health satisfaction	0.02*	-0.01	-0.02	-0.02	-0.01	-0.00	0.01	0.03
Life satisfaction	0.01*	0.01*	-0.00	0.02	-0.00	0.01	-0.03**	-0.01
Employment level								
Employed	0.01	-0.04	0.03	-0.04	0.03	0.01	0.05	-0.02
Retired	0.02	-0.02	0.09	0.07	0.19*	-0.08	-0.03	-0.07
n	2105	422	422	422	422	422	309	309
LR Chi2	348.39	10.26	6.34	16.74	9.02	10.37	18.23	8.01
Pseudo R2	0.16	0.05	0.01	0.04	0.02	0.02	0.02	0.02
Log Likelihood	-899.61	-97.88	-242.03	-187.55	-281.83	-267.60	-210.16	-163.32

*= significant at 0.05; **= significant at 0.1; Health state, v value and duration have only been analysed for type I questions

Figure 1: Basic binary choice question format used in the survey

Question Types I-V

	Health scenario A	**Health scenario B**
	Person (P) lives for (L) years in full health followed by duration (T) years in state (H) with satisfaction level (S) and then dies	Person (P) lives in full health for (L+VT) years and then dies
Which scenario do you think is better?		

Question Type VI

	Health scenario A	**Health scenario B**
	You live for (L) years in full health followed by duration (T) years in state 55555 and then die	You die immediately
Which scenario do you think is better?		

Question Type VII

	Health scenario A	**Health scenario B**
	You live in health state (H) for T years and then die	You live in health state (H) for T years and then die
Which scenario do you think is better?		

Figure 2: Self reported health status and health and life satisfaction