Health state valuation questions: Does mode of administration matter?

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

1. Health Economics and Decision Science, School of Health and Related Research, University of Sheffield, Regent Court, Sheffield, S1 4DA, UK
2. Department of Economics, University of Sheffield, Sheffield, S1 4DT
3. Office of Health Economics, 12 Whitehall, London, SW1A 2DY
4. Centre for Health Evaluation and Outcome Sciences, St Paul's Hospital, Vancouver, BC, V6Z 1Y6
5. Health Economics Research Group, Brunel University, UB8 3PH

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

Aims The first aim of this paper is to give an overview of the methodological issues addressed in an on-going MRC funded study which aims to inform the re-evaluation of the EQ-5D tariff (PRET). The second aim is to report a study within this project investigating whether responses to a range of health state valuation questions differ by administration mode (online vs. face-to-face). ***Methods*** PRET includes questions investigating a range of methodological factors relating to health state valuation and these are described in detail. For the second aim, two identical surveys including binary choice health state valuation questions were administered in online and Computer Assisted Personal Interview (CAPI) settings. Data collected included demographics, self reported health status, health and life satisfaction and EQ-5D-5L scores, and responses to the valuation questions. To investigate differences, descriptive and regression analyses were conducted. ***Results of aim 2*** Overall, 422 respondents completed a survey (221 online; 201 CAPI). There were no overall age or gender differences. However, online respondents were more likely to be educated to a higher level and employment status differed. CAPI respondents reported significantly better health status, health/life satisfaction and EQ-5D-5L scores. CAPI also took significantly longer to complete. However there was no effect of administration mode on responses to the valuation questions, and this was replicated when controlling for demographic differences. ***Conclusions*** The responses to the experimental binary choice questions were not significantly different by mode of administration, and this is not significant in multivariate probit analyses explaining the binary choices.

Acknowledgements
This project is funded by the MRC Methodology programme (ref.G0901500). BM is partly funded by the MRC. We are grateful to all respondents who took part. Research ethics approval has been given by the ethics committee at the School of Health and Related Research in line with University of Sheffield research governance and ethics requirements. The usual disclaimer applies.

1. Background: the PRET project

Resources are limited and need to be allocated efficiently. The health care sector is no exception. The National Institute for Health and Clinical Excellence (NICE) was set up to help make better health care resource allocation decisions. NICE bases its recommendations on cost effectiveness analyses with the Quality Adjusted Life Year (QALY) as the outcome measure. The EQ-5D (Brooks, 1996), is the preferred instrument to use when quantifying the health related quality of life (HRQOL) impact of medical interventions (NICE, 2008). The EQ-5D investigates HRQOL across five dimensions (mobility, self care, usual activities, pain/discomfort and anxiety/depression) each with three response levels (no, some or extreme problems).

The EQ-5D has an associated population value set, or "tariff" score. The current tariff is based on the Measurement and Valuation of Health (MVH) study, carried out in 1994 and published in 1997 (Dolan, 1997). The study used face-to-face interviews of a representative sample of the UK general population. A selection of hypothetical EQ-5D states were assessed using the time trade-off (TTO) method. The results were modelled using regression to provide a population value set, which in effect is a tariff of HRQOL weights for each of the 243 EQ-5D health states.

The UK EQ-5D value sets are used not just by NICE, but also as the basis for economic evaluation by other decision makers and researchers both in the UK and elsewhere. They are also used in a range of further applications, including population health surveys (eg. the Health Survey for England); burden of disease studies; hospital inpatient surveys and, most recently, the NHS PROMs initiative (Browne et al 2007).

In the 17 years since the MVH study, there have been developments that have lead to the need for a re-evaluation of the EQ-5D tariff:

- People may not have the same preferences as they did 17 years ago;
- Change in demography may mean that although individual preferences may not have changed, the composition of people across the country has changed, so that average preferences may have changed;
- Recognition of the shortcomings of the MVH TTO design, in particular in the context of observations worse than dead;
- New advances in methods for valuing health states other than TTO, such as discrete choice experiments (DCE);
- New advances in the mode of valuation, other than face-to-face interviews; and
- The development of a revised version of the EQ-5D, with 5 levels rather than 3 (EQ-5D-5L).

In order for NICE to make the most appropriate decisions, the EQ-5D UK population value set needs to be up to date and based on the latest understanding of health state preferences. A study with the aim of re-evaluating the EQ-5D tariff will be carried out in the near future. The MRC funded study "**P**reparatory study for the **R**e-evaluation of the **E**Q-5D **T**ariff" (PRET) is a methodological study that aims to contribute to the re-evaluation of the EQ-5D population value set, by exploring a range of methodological issues associated with health state valuations.

PRET has four stages. In Stage 1, a large scale online survey is carried out to explore a series of methodological issues. In Stage 2, a part of the online survey is carried out in a face-to-face environment using Computer Assisted Personal Interviewing (CAPI) technology on a laptop. Stages 3

and 4 consist of more detailed interviews where selected methodological issues are examined in more detail.

The purpose of this paper is twofold. The first is to give an overview of the methodological issues addressed in PRET. The second aim is to report a study investigating the similarities and differences between the online and CAPI modes of administration. This draws on data from Stages 1 and 2 of PRET. Section 2 of this paper includes an overview of the methodological issues addressed in PRET, and section 3 presents the study design of Stages 1 and 2. Section 4 presents the methods of the data collection, recruitment, and analysis. Sections 5 and 6 are for the results and the discussion.

2. Methodological issues addressed by the PRET project

The design of a valuation study for any health state classification instrument will need to take the following issues into account:

(1) Whose values to obtain?
(2) Which health state classification system?
(3) What mode of administration?
(4) What method of valuation?
(5) How many, and which hypothetical health states to value?
(6) How long should each hypothetical state last?

Each is discussed in more detail below.

(1) Whose values?

The current MVH value set is based on general population values. General population values of hypothetical health states may differ from the way patients value hypothetical states or their own current state, and there has been debate about which values should be used (Brazier et al, 2005). PRET does not have the capacity to compare patient and general public values, and only members of the UK general population are included. However, a recent study has demonstrated that if the general public can be informed about the extent to which it is possible for patients to be satisfied with their condition, the discrepancy in values may diminish (McTaggart-Cowan et al, 2009). PRET examines this further by introducing an element of health satisfaction, so that the way in which patients feel about the state of health can be captured.

There is also a normative element to this debate, concerning whether general public values ought to be used over patient values. The use of general public values is typically justified with reference to the non-welfarist argument which states that as the values are used in decision making in a publicly funded health care system, they should come from people as informed citizens, not from people as consumers (see for example Tsuchiya and Miyamoto, 2009). The traditional approach to health state valuation, and that used for the current MVH value set, has been to obtain valuations through asking respondents to imagine themselves in the health state of interest. If an informed citizen perspective is taken a different framing of the TTO question may be required to reflect that the respondent is valuing health states on behalf of other members of society. However, it is unclear what impact an alternative perspective will have on values. PRET investigates this by comparing responses using the standard individual perspective with two alternatives reflecting the citizen

approach. In addition, background characteristics questions allow for an examination of the respondent's illness experience and satisfaction with their own health.

(2) Which health state classification system?

There are two issues here. The first is which version of EQ-5D to use. The version of EQ-5D that is most commonly in use is the three level version (Brooks, 1996). This is the version of EQ-5D that the MVH population value set relates to. However, the EuroQol Group has recently released the 5-level version of EQ-5D (Herdman et al, forthcoming) and the next UK valuation study of the EQ-5D is likely to be focussed on this instrument. Therefore, PRET uses EQ-5D-5L health states where relevant.

A second issue is whether or not the five dimensions of EQ-5D cover all relevant aspects to be taken into account when health state valuation exercises are conducted. There is a limit to the number of items respondents can cope with in a health state valuation exercise, so if the aim is to retain a unique generic instrument applicable to all conditions, then any further information to be incorporated also needs to be fairly generic. (On the other hand, if different condition-specific add-ons are used, then each additional item need not be generic.) PRET looks at the implications of introducing an element of health satisfaction alongside the standard EQ-5D health state descriptive system.

(3) What mode of administration?

The current MVH TTO value set is based on face-to-face interviews. Whilst this is a method that results in high quality data, it also very expensive method. When the MVH study was carried out, there were two more alternatives available: postal questionnaire and telephone interview (with or without a pre-posted questionnaire). While these two modes are much less costly than face-to-face interviews, they are usually regarded as resulting in lower quality data.

However, over the past decade there have been major advances in communication technology, and one attractive mode of survey administration is via the internet. These surveys use online panels, including a pool of potential respondents with registered background characteristics. One of the purposes of this paper is to report on a head to head comparison between an online administration (in Stage 1) and a CAPI administration (in Stage 2) of an otherwise identical survey.

(4) What method of valuation?

The current MVH value set is based on TTO (Gudex, 1994). This TTO protocol is known to have inherent problems, in particular, regarding the procedure used to value states worse than dead. It is not only different from but also incommensurable with the procedure used for states better than dead (Tilling et al, 2010). While in the UK value set the average values for two thirds of states are positive, a large number of states have individual observations that are negative, and this can distort the average values. An alternative TTO protocol called the 'lead time TTO' has been devised (Robinson and Spencer, 2006; Devlin et al, 2010). This processes all states in the same way, regardless of whether they are better or worse than dead, by adding a set number of years in full health preceding each of the 'lives' being valued in the time trade off exercise. Further analysis is required to identify the optimal length of this lead time, and one of the objectives of PRET is to provide evidence on this issue. Moreover, one concern is that if the value of a health state depends

on its timing and on a preceding health state, then the addition of lead time may distort the TTO value. PRET includes a comparison of the MVH TTO and the lead time TTO using binary choice questions as explained in section 3 below.

In addition to TTO, there is a growing interest in the application of DCE in health state valuation. One advantage of the DCE is that because individuals are not interrogated until they reach a point of indifference (as in iterative TTO), but only asked to give ordinal preferences over pairwise choices, it is arguably less cognitively demanding than such methods. On the other hand, the well-known problem with the DCE has been that there has been no satisfactory method of combining the dimensions of health with survival and duration. However, a method has been developed that interprets DCE data as a TTO exercise (Bansback, et al, 2010). The method includes duration as one of the DCE attributes and estimates a regression model with an interaction term between health state and duration. The coefficients are used to calculate the value of health states by solving the equivalence relationship for a binary choice situation between, on the one hand, living in a given health state for a specific duration of time and, on the other, living in full health for a shorter duration. This is equivalent to the indifference point in TTO. Not only does this potentially solve one of the key the problems faced by DCE, it also potentially solves the issue of observations worse than dead without recourse to the lead time structure (see Bansback et al., 2010). PRET explores this approach further.

(5) Which hypothetical health states to value?

The original 3-level EQ-5D has 243 possible states. The current MVH TTO value set is based on direct valuations of 45 of these. However, the introduction of EQ-5D-5L means that there are now 3,125 possible health states to model. Findings from PRET may be used as prior information to assist in the selection of states and design of the re-evaluation study.

(6) How long should each hypothetical state last?

The current MVH TTO value set is based on participants being asked to imagine each health state lasting for a duration of 10-years. However, the MVH also estimated TTO tariffs for different durations. This was because there was a concern that the tariff values may be a function of the duration of the health state. There are four related issues, all of which are also relevant to DCE (see Tsuchiya and Dolan, 2005). One is whether or not 'constant proportional time trade off' holds so that the utility associated with a marginal survival in a given health state remains constant regardless of the health state or the duration. It has been argued that for very severe states, there may be a 'maximal endurable time' limit, beyond which the marginal benefit of survival diminishes (Stalmeier et al. 2007). The second issue is whether or not respondents use a positive temporal discount rate when valuing hypothetical health scenarios. The third is the impact of life stage concerns in health state valuations. If the duration of the state is too long, then the scenarios will not be credible for older respondents and vice versa. Furthermore, depending on the duration, people may be thinking about life stage events rather than about the trade off between longevity and quality of life.

The final issue is whether or not 10-years is the most relevant duration of health states for NICE decision making. If the above issues mean that the value of a state is a function of its duration, and if most NICE decision making involve states that last for much shorter durations, then the re-

evaluation of the EQ-5D should not be based on scenarios with a 10-year duration. Therefore, PRET examines the impact of duration on health state preferences.

Thus, to summarise, there are a number of issues that need to be addressed before the re-evaluation of the EQ-5D value set can take place. PRET examines these and, based on the findings, aims to recommend a study protocol to inform the re-evaluation of the EQ-5D tariff.

3. The study design used in Stages 1 and 2

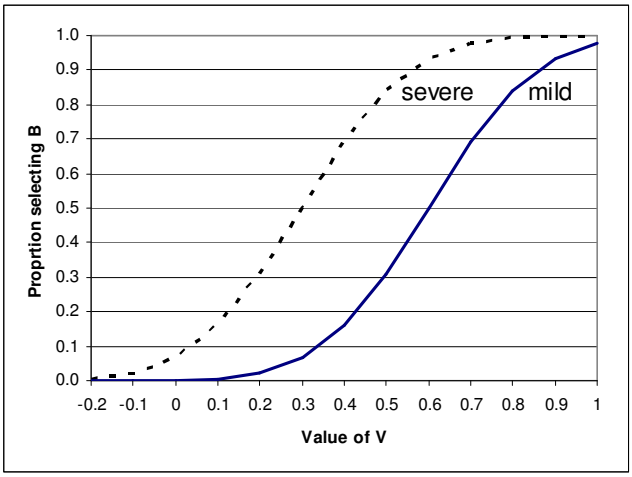
3.1 The survey question format and Types of questions in Stages 1 and 2

The survey questions in Stages 1 and 2 take the format of binary choice. A single response to a binary choice question cannot identify the level of HRQOL an individual feels is right for a given health state. However, by examining the distribution of responses of multiple respondents across different binary choice questions, the relevance of key parameters can be identified.

The most "basic" binary choice question used in PRET is as follows:

> **[Scenario A]**: you will live in health state **H** for 10 years and die
> **[Scenario B]**: you will live in full health for (**V** x 10) years and die
>
> Which of the two scenarios do you think is better?

The figure on the right illustrates a highly stylised example of two hypothetical states, "severe" and "mild". Along the horizontal axis is the value of V with 0 for dead and 1 for full health. Along the vertical axis is the proportion of people. The upward sloping curves indicate that, as V increases, the proportion of people who think the given health state is no better than V will increase. Now, suppose V is 0.6. If the state H in the example above is the severe state, then around 90% of people think it is no better than 0.6 and thus choose Scenario B. But if state H is the mild state, then around 50% will think it is no better than 0.6 and thus choose Scenario B.



All binary choice scenarios will include information on a health state and the length of time lived in the state, followed by death. These binary questions could be part of a DCE design. At the same time, they are a snippet of a TTO procedure; the typical TTO exercise involves changing V until the respondent is indifferent between the two scenarios. In fact, TTO can be interpreted as a special case of DCE, where Scenario B always involves full health (Bansback et al, 2010).

There are seven *Types* of binary choice questions used in this paper and these are summarised in Table 1. Question types include one or more of the following key parameters:

- EQ-5D-5L **states** (**H**): see section 3.3 below

- **Duration** in years (**T**) in state H. PRET uses 10 weeks, 1 year, 5 years and 10 years.
- **Lead time** stretches (**L**) in full health including zero, 10 weeks, 1 year, 5 years and 10 years
- Person **perspective** (**P**) that the TTO applies to. PRET uses "you", "somebody else like you", "somebody else".
- Levels of **satisfaction** with one's own health or life (**S**). PRET uses low, medium and high and also "learnt to live with health state".

Not all of the question Types mention all parameters. Each question type is explained below.

3.2 *The assumptions tested using Type I to Type V questions*

The first five question Types are based on TTO, and present two scenarios:

- [A]: Person **P** lives **L** in full health followed by **T** in state **H** with satisfaction **S** then dies
- [B]: Person **P** lives (**L** + **VT**) in full health then dies

Scenarios in Type I use the "you" perspective, and do not mention lead time (L) or satisfaction (S). Scenarios in Type II use the "other" perspective, but are otherwise identical to Type I. Scenarios in Types III and IV mention lead time (L): Type III use the "you" perspective and Type IV the "other" perspective. Scenarios in Type V include satisfaction (S). By comparing the responses to the first five question Types, the following assumptions are explored.

<u>Assumption 1: health state values are independent of duration</u>

Type I questions are used. If the assumption holds, then for a given combination of state H and value V, the distribution of respondents between the two scenarios should not be affected by duration T. Therefore, when changing from one duration to another in Scenario A the cumulative curve for the state should remain in the same position, and the proportion of people choosing each Scenario at a given V should be unaffected.

<u>Assumption 2: health state values are independent of person perspectives</u>

Types I and II questions are used. If the assumption holds, then for a given combination of state H and value V, the distribution of respondents should not be affected by person perspective P.

<u>Assumption 3: health state values are independent of lead time</u>

Types I and III questions are used. If the assumption holds, then for a given combination of state H and value V, the distribution of respondents should not be affected by lead time L.

<u>Assumption 4: the values of others' health are independent of when health events take place</u>

Types II and IV questions are matched so that they are identical except for the timing of health events. The exercise is not affected by life stage considerations that inevitably affect time preference exercises using the "you" perspective. If the assumption holds, then for a given combination of state H, value V, and person perspective P, the distribution of respondents should not be affected by the timing of health events, represented by lead time L.

<u>Assumption 5: the values of others' health are independent of satisfaction in the state</u>

Type V questions are used. If the assumption holds, then for a given combination of state and value V, the distribution of respondents should not be affected by satisfaction S.

3.3 *The hypothetical health states used in Type I to Type V questions*

Questions Type I to V will use the following five health states:

- "Slight problems walking about" (EQ-5D-5L state 21111)
- "Slight pain" (11121)
- "Unable to walk about" (51111)
- "Extreme pain" (11151)
- "Extreme depression" (11115)

Only one dimension in any health state represents a health problem, and therefore, these states are easy to imagine. In addition, they cover different aspects of health, and thus enable a test of the key assumptions across different kinds of health problem. Two sets of V were used: 0.8 and 0.9 for the mild states ("slight") and 0.4 and 0.6 for the severe states ("unable"/"extreme"). These values were chosen based on the MVH tariff values for the five comparable health states in EQ-5D-3L.

3.4 *Type VI questions to test the sufficiency of lead time under very poor health*

Type VI binary choice questions look like this:

- [A]: You will live **L** in full health followed by **T** in state 55555 then die
- [B]: You will die immediately

Developmental work for lead time TTO has indicated that some respondents associate very poor states with extreme negative values. Therefore they will 'use up' or exhaust all their lead time (Devlin et al, forthcoming). When this happens, no TTO value can be inferred. While at least some of these may reflect a genuine quantitative preference, others may be a qualitative indication that the state is extremely poor. Type VI questions aim to map the proportion of respondents who exhaust lead time at various combinations of duration T and lead time L to gauge the proportion of respondents who may be giving a qualitative preference for very poor states. The worst possible EQ-5D-5L state 55555 (extreme problems in all five dimensions) is used for all scenario combinations.

3.5 *Type VII questions for informing the selection of states for DCE*

Type VII questions are in effect a small scale DCE study and take the following form:

- [A]: You live duration $\mathbf{T_A}$ in state $\mathbf{H_A}$ then die
- [B]: You live duration $\mathbf{T_B}$ in state $\mathbf{H_B}$ then die

Both scenarios here consist of "you" living in a particular EQ-5D-5L state H_i for a specified duration T_i followed by death (i=A,B). The aim of Type VII questions is to generate data that can be used as prior information to guide the design of an efficient set of hypothetical health states to be used in the re-evaluation of EQ-5D-5L, and this will be reported elsewhere.

3.6 *The allocation of questions to questionnaire versions*

Different *Types* of questions are presented across three *Modules*:

- Module 1: Five Type I questions
- Module 2: Five questions specific to the questionnaire version (one of Type II to Type VI)
- Module 3: Two Type VII questions

Therefore each respondent will be presented with 12 *binary choice questions*. The five health states will be used only once each in Modules 1 and 2, combined with a different duration T. There are 15 versions of the online questionnaire. In 14 of these, Module 2 consists of five binary choice questions from one of Types II, III, IV, V or VI, so that respondents given these versions face three question types each. However, in version 15, Module 2 consists of one question each from Types II, III, IV, V or VI. Therefore respondents allocated to version 15 face all seven question Types. All respondents in the face-to-face CAPI sample are given version 15. Thus, the comparison between online and CAPI administrations reported in this paper involves all seven question Types (see table 2 for a description of the 12 questions used in version 15).

Furthermore, there are 60 *sub-versions* (each of the 15 versions has four sub-versions) for Module 3. EQ-5D-5L has 3,125 possible health states, and combining this with three levels of duration amounts to 9,375 possible DCE scenarios. Of these, 240 were selected and paired using a D-Optimal algorithm which selects states using the full factorial design as the starting point. Two of the 120 pairs are allocated to each of the 60 sub versions, and three duration levels (1 year, 5 years and 10 years) are used for these questions.

3.7 *The mode of administration in Stages 1 and 2*

Interest in using internet surveys to collect data has grown in recent years, and methods of utilising the web to conduct health state valuation surveys are starting to be investigated, moving the methodologies on from the traditional face-to-face administration. For valuation studies, both face-to-face interviews and online methods have clear advantages and disadvantages. Face-to-face interviews provide high quality and complete data, but they are expensive and time consuming to conduct. The advantages of online surveys are that complex routing (or branching) of questionnaires is possible; question ordering can be easily randomised; the time taken for each question (or groups of questions) can be logged; there is no process of data entry and associated errors as this is done automatically; a wide range of background characteristics of non-respondents can be obtained; large samples can be achieved in a short time; and the sampling frame can be flexible. However there are also concerns around the representativeness of the sample in terms of unobserved characteristics; the motive of participation; the level of non-response (Eysenbach, 2005); data quality (Bowling, 2005); and whether respondents are genuinely engaged in the task . The use of CAPI means some of the advantages of online surveys can be exploited in face-to-face environments, but not all.

Limited work has compared the online and CAPI administration of health state valuation exercises. Norman et al. (2010) have compared the online and face-to-face 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. This study aims to compare responses to the binary choice valuation across online and CAPI modes of administration.

4. Methods

4.1 The survey

The surveys administered contained identical questions in the same order, with the only difference being the mode of administration. Each survey begins by providing a brief background and explaining the purpose of the survey. This is followed by an informed consent page. Consenting respondents are then asked a series of background characteristic questions. If the respondent is of an age or gender group with a complete quota, they do not continue from this point. Continuing respondents complete questions about health status, health and life satisfaction and the EQ-5D-5L before proceeding to the experimental questions. After completion there is an opportunity for respondents to give free text comments on the survey.

4.2 Recruitment and the sample

In order to achieve a comparison of the two modes of administration as they would happen in the real world, the two samples were recruited separately by following procedures that would be employed in typical surveys. Both aimed for an achieved sample of 200.

For the Stage 1 online survey, respondents were sourced from an existing internet panel, and were selected following set quotas for age and gender to ensure that the sample was representative of the UK general population. Invitations were sent out by e-mail, and respondents clicked a link to access the survey, read the project information, consented to take part, and answered demographic screening questions. 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 went through the survey in less than the minimum imposed time limit of 5 minutes. The online survey was hosted by a market research company separate from the recruiting agency.

For the Stage 2 face-to-face interviews, recruitment was carried out by the company hosting the online survey. Participants were recruited by knocking on one in every ten doors in selected postcodes in five UK areas. The interviewer explained the project and sought consent to take part. The survey was presented to respondents on a laptop, and the interviewer read out all of the questions and recorded the response given. This was done in a one-to-one setting. Again, participants were selected following set quotas for age and gender. A minimum completion time of 5 minutes was imposed, and participants were able to stop the survey at any time.

4.3 The analysis

Version 15 of the Stage 1 online survey and the entire Stage 2 CAPI-generated data are used for the analysis. Background characteristics, self reported health, and time taken to complete the survey are compared across the two samples using chi squared and ANOVA. This is followed by a comparison of the proportion of respondents who choose scenario B, by sample, and by question Type. Significance is indicated by p values < 0.05. Probit regressions are used to explore the determinants of the propensity to choose scenario B for each question:

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

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 H, T, L, P, S as earlier defined, and the function $\Phi(.)$ is the distribution function of the standard normal distribution (Greene, 1998; Alexandre and French, 2004). The advantage of reporting marginal effects is that 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 22%. For the regressions, significance levels of both <0.05 and <0.1 are used.

5. Results

5.1 *Respondent characteristics and self reported health*

In total, 422 respondents completed either the relevant online survey or the CAPI version. For the online survey 2326 panel members were invited to take part and 487 (20.1%) accessed the survey. Of these 266 (11% of those invited; 54% of those accessing the survey) were screened out (due to quota), 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 5 minutes or more. There were no demographic 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. 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.

The characteristics of the samples are displayed in Table 3. Overall, the CAPI sample was more comparable to the UK general population (ONS, 2001). There were no significant differences between the groups by age and gender, but a number of demographic variables significantly differed between the samples. These included employment status, 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 terms of completion time, the CAPI sample took significantly longer to complete the overall survey, and also the experimental questions included in Modules 1 and 2. There were no differences between the samples for the time taken to complete the Module 3 questions.

Responses to the self report health status and health and life satisfaction questions are displayed in Figures 1-3. The CAPI sample are significantly more likely to report better health (p = 0.002) and higher levels of health (p = 0.000) and life satisfaction (p = 0.000). The EQ-5D indices ($F(1,409)$ = 16.51, p = 0.000) and the dimensions responses also differ significantly by mode of administration with the exception for mobility, with the CAPI group reporting better health (see Table 4).

5.2 *Binary choice valuation questions*

The proportion of the sample choosing scenario B (i.e. choosing to live for a shorter duration in full health) did not significantly differ by administration mode for any of the seven binary choice question Types. This was irrespective of the health state, duration or length of time in full health

presented in the scenario (see Table 5). Probit regressions reveal that a range of demographic and question design related 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. This can be seen in Table 6, which displays the coefficients a reduced model after collinear variables were dropped. 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 it is that 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 and those with higher levels of life satisfaction are more likely to choose to live in full health. For Type IV, females are again more likely to choose scenario B, and for Type V, males and respondents who are retired are 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.

6. Discussion and conclusion

When the MVH study was conducted and the EQ-5D value set for the UK was estimated, face-to-face interviews were seen as the best way to administer health state valuation exercises. However, in the years since, there have been dramatic advances in communication technology, and interest in the use of online surveys is growing. On the other hand, there are also issues regarding the quality of data generated in this way. This paper reports on a comparison between an identical set of binary choice questions designed for health state valuations conducted in online and face-to-face environments. The results show differences between the modes in terms of the characteristics of the sample, but not in the responses to the binary choice health state valuation questions.

The two samples were recruited against quotas set in terms of age and sex, and thus they do not differ in terms of these characteristics. However, the two samples are statistically significantly different in terms of observable characteristics such as education, employment status, health, and health/life satisfaction. In theory, it is possible to make the two samples agree in terms of any observable characteristic: it is a matter of screening potential respondents before they enter the main survey. However, even with highly selective screening, the possibility will remain that the two samples differ in terms of further unobserved characteristics. Because the face-to-face interview has for a long time been regarded as the ideal mode of survey administration, it is easy to conclude that it is the online survey which is recruiting a more idiosyncratic sample, in effect implying that the kind of people who are registered to internet survey panels cannot be representative of the wider public. At the same time, however, it should also be noted that not everybody is equally likely to take part in a face-to-face interview: they may not be at home at the kinds of time when the interviewer knocks on their door, or may not consent to allow a stranger interviewer to enter their home. Typically, characteristics of non-responders to interviews are not available. One advantage on the other hand of online surveys is that certain characteristics of non responders may be available, allowing an analysis of these issues.

A completely pragmatic approach is to look at the data that are generated, firstly on time taken, and secondly the actual responses to the experimental questions. Looking at the data on time taken may allow insights regarding engagement. If an online respondent completes the survey too quickly or too slowly, this may suggest that they are not fully engaged. On the other hand, with a CAPI respondent, 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 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. However, when the actual responses are compared, they are not statistically significantly different from each other, with high p-values consistently (Table 4). This finding was robust and was not influenced by the severity or duration of the state (as the findings are consistent irrespective of the proportion of respondents choosing scenario B), or by the different composition of the background characteristics of the two samples.

In a study comparing iterative TTO using the EQ-5D-3L across modes of administration, Norman et al, (2010) found differences between online and CAPI responses, and suggested that this was due to the iterative nature of the process. A particular concern for iterative exercises in an online environment is that if respondents intend to get through the questions quickly, there is an incentive to accept the first trade off offered, without going through the process to reach indifference. Therefore any successful online version of the TTO is unlikely to be a simple transplant of an existing interview-based iterative protocol to an online environment. Our study did not use an iterative process, but rather binary choice 'snapshots' of TTO, which is equally amenable to online and CAPI administration. And as such, provides a fairer comparison of the two modes of administration. Furthermore, our study found that the mode of administration did not affect response to the Type VII DCE questions. In other words, where a design that is equally suited to online and CAPI administration is used, the mode of administration cannot be said to impact on the results.

Furthermore, the financial and time costs of the surveys are worth considering. The cost of any survey has a fixed element and is not completely proportionate to sample size. However, very roughly, the cost of an online survey using a commercial internet panel is likely to be in the range of £2-15 per respondent, and a sample of 3000 can be achieved in a couple of weeks. However, face-to-face interviews can be 10 times more costly per respondent and may take up to 10 times longer to recruit sufficient numbers of participants. Therefore, when the survey design is amenable to online administration, the incremental cost effectiveness of conducting interview surveys must be examined.

This paper has presented an overview of the methodological issues regarding health state valuation studies that are explored in PRET, and discussed the findings from a head to head comparison of online and CAPI administrations of binary choice questions. The two administrations have different advantages and disadvantages, and resulted in two samples with statistically significant differences in some background characteristics. However, their responses to the experimental binary choice questions were not significantly different, and the mode of administration is not significant in multivariate probit analyses explaining the binary choices.

[main text: 6442wds]

References

Alexandre PK, French MT (2004). 'Further Evidence on the Labor Market Effects of Addiction: Chronic Drug Use and Employment in Metropolitan Miami', *Contemporary Economic Policy*, 22: 382-393.

Bansback N, Brazier J, Tsuchiya A, Anis A (2010), *Using a discrete choice experiment to estimate societal health state utility values*, HEDS Discussion Paper, ScHARR, University of Sheffield

Bliemer MCJ, Rose JM (2006) *Designing Stated Choice Experiments: State-of-the-art*, paper presented at the 11th International Conference on Travel Behaviour Research, Kyoto

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

Brazier J, Akehurst R, Brennan A, al (2005), Should patients have a greater role in valuating health states? *Applied Health Economics and Health Policy*, 4: 201-208

Browne J, Jamieson L, Lawsey J, van der Meulen J, Black N, Cairns J, Lamping D, Smith S, Copley L, Horrockes J (2007) *Patient Reported Outcome measures (PROMs) in elective surgery*. Report to the Department of Health

Brooks R (1996). EuroQol: The current state of play. *Health Policy*, 37:53-72

Devlin N, Tsuchiya A, Buckingham K, Tilling C (forthcoming), A Uniform Time Trade Off Method for States Better and Worse than Dead: Feasibility Study of the 'Lead Time' Approach, *Health Economics*

Dolan P (1997) Modelling valuations for EuroQol health states. *Medical Care*, 35(11): 1095-1108

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

Greene, W.H. (2000). *Econometric Analysis*, New Jersey: Prentice Hall.

Gudex C ed (1994) *Time Trade-Off User Manual: Props and Self-Completion Methods*: Centre for Health Economics, University of York

Herdman M, Gudex C, Lloyd A, Kind P, Parkin D, Bonsel G, Badia X (forthcoming), Development and preliminary testing of the new five-level version of EQ-5D (EQ-5D-5L), *Quality of Life Research*

McTaggart-Cowan H, O'Cathain A, Tsuchiya A, Brazier J. (2009) *A Qualitative Study Exploring the General Population's Perception of Rheumatoid Arthritis after being Informed about Disease Adaptation*. Health Economics and Decision Sciences Discussion Paper 09/02

National Institute for Health and Clinical Excellence (2008), *Guide to the Methods of Technology Appraisal*

Robinson A, Spencer A (2006) Exploring challenges to TTO utilities: valuing states worse than dead. *Health Economics*, 15: 393–402.

Stalmeier PF, Lamers LM, Busschbach JJ, Krabbe PF (2007). On the assessment of preferences for health and duration: maximal endurable time and better than dead preferences. *Medical Care*, 45(9) 835-841.

Tilling C, Devlin N, Tsuchiya A, Buckingham K (2010). Protocols for TTO Valuations of Health States Worse than Dead: A literature review and framework for systematic analysis, *Medical Decision Making*, 30:610–619

Tsuchiya A, Dolan P (2005), The QALY model and individual preferences for health states and health profiles over time: A systematic review of the literature, *Medical Decision Making*, 25(4)460-467

Tsuchiya A, Miyamoto J (2009), Chapter 20 Social choice in health and health care, in Anand P, Puppe C, Pattanaik P eds, *Handbook of Rational and Social Choice*, Oxford University Press

Tables and figures

Table 1: The seven *Types* of questions used in Stages 1 and 2

Question type / Parameter	I	II	III	IV	V	VI	VII
State of health (H)	√	√	√	√	√	√	√
Duration in full health (T)	√	√	√	√	√	√	n/m
Duration in H	√	√	√	√	√	√	√
Lead time (L)	n/m (*)	n/m	√	√	n/m	√	n/m
Person/perspective (P)	you	other	you	other	you	you	you
Satisfaction (S)	n/m	n/m	n/m	n/m	√	n/m	n/m

(*) n/m means not mentioned in the scenario

Table 2: 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

Table 3: Sample characteristics

Characteristic	Overall (%)	Online (%)	CAPI (%)	Significance
Age				
Mean (SD)	41.49 (13.96)	41.56 (14.38)	41.41 (13.52)	*P* = 0.913
Range	18-66			
Age category (n,%)				*P* = 0.233
18-24	64 (15.2)	34 (15.4)	30 (15.0)	
25-34	97 (23.0)	51 (23.1)	46 (22.9)	
35-44	85 (20.1)	36 (16.29)	49 (24.4)	
45-54	86 (20.4)	46 (21.8)	40 (19.9)	
55+	90 (21.3)	54 (24.4)	36 (17.9)	
Male (n,%)	201 (47.6)	102 (46.2)	99 (49.3)	*P* = 0.524
Employment (n,%)				*P* = 0.009
In employment	245 (58.1)	128 (57.9)	117 (58.2)	
Retired	41 (9.7)	17 (7.7)	24 (11.9)	
Homemaker	34 (8.1)	10 (4.5)	24 (11.9)	
Student	36 (8.5)	23 (10.4)	13 (6.5)	
Seeking work	16 (3.8)	10 (4.5)	6 (3.0)	
Unemployed	18 (4.3)	9 (4.1)	9 (4.5)	
Long term sick	25 (5.9)	18 (8.1)	4 (2.0)	
Other	7 (1.7)	6 (2.7)	1 (0.5)	
Marital status (n,%)				*P* = 0.039
Married/partner	236 (55.9)	111 (50.2)	125 (62.2)	
Single	135 (32.0)	78 (35.3)	57 (28.4)	
Education continued after minimum age (n,%)	292 (69.2)	174 (78.7)	118 (58.7)	*P* = 0.000
Educated to degree level (n,%)	136 (29.9)	90 (40.7)	46 (22.9)	*P* = 0.032
Time taken to complete (M (sd) minutes)				
Overall	9.88 (4.6)	8.64 (3.84)	11.26 (4.99)	*P* = 0.000
Module 1	1.27 (0.76)	1.07 (0.77)	1.49 (0.70)	*P* = 0.000
Module 2	1.92 (1.33)	1.80 (1.63)	2.06 (0.89)	*P* = 0.045
Module 3	1.28 (0.99)	1.20 (1.11)	1.36 (0.84)	*P* = 0.088

Table 4: EQ-5D-5L responses by mode of administration

	Online (%)	CAPI (%)	Significance
EQ-5D index score	0.720 (0.33)	0.843 (0.28)	*P* = 0.001
Mobility			*P* = 0.055
1	160 (72.7)	168 (83.4)	
2	29 (13.2)	18 (9.0)	
3	18 (8.2)	7 (3.5)	
4	11 (5.0)	8 (4.0)	
5	2 (1.0)	0 (0)	
Self care			*P* = 0.029
1	193 (88.9)	194 (96.5)	
2	10 (4.6)	2 (1.0)	
3	11 (5.1)	4 (2.0)	
4	3 (1.4)	1 (0.5)	
5	0 (0)	0 (0)	
Usual activities			*P* = 0.001
1	148 (67.6)	172 (85.6)	
2	37 (16.9)	14 (7.0)	
3	20 (9.1)	11 (5.5)	
4	13 (5.9)	3 (1.5)	
5	1 (0.5)	1 (0.5)	
Pain/discomfort			*P* = 0.001
1	98 (44.7)	132 (66.0)	
2	70 (32.0)	40 (20.0)	
3	37 (16.9)	18 (9.0)	
4	9 (4.1)	6 (3.0)	
5	5 (2.3)	4 (2.0)	
Anxiety/depression			*P* = 0.001
1	120 (54.8)	154 (77.0)	
2	52 (23.7)	24 (12.0)	
3	31 (14.2)	18 (9.0)	
4	8 (3.7)	2 (1.0)	
5	8 (3.7)	2 (1.0)	

Table 5: Proportion of respondents choosing Scenario B in different binary choice questions

Question	Type	Online (%)	CAPI (%)	Sig
1	I	67.0	68.2	0.794
2	I	54.8	58.7	0.413
3	I	81.9	81.6	0.935
4	I	98.2	98.5	0.799
5	I	91.9	91.5	0.907
6	II	92.8	94.0	0.601
7	III	71.0	75.6	0.288
8	IV	81.9	83.1	0.749
9	V	56.6	60.7	0.389
10	VI	65.6	64.6	0.841
11	VII	49.1	49.8	0.910
12	VII	77.8	76.6	0.817

In questions 1-9, Scenario B represents living in full health for a shorter duration.
In question 10, Scenario B represents immediate death.
In questions 11-12, Scenario B is a 5 level EQ-5D-5L health state with associated duration.

Table 6: Probit regression 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 has only be analysed for type I questions

Figure 1: Self reported health status

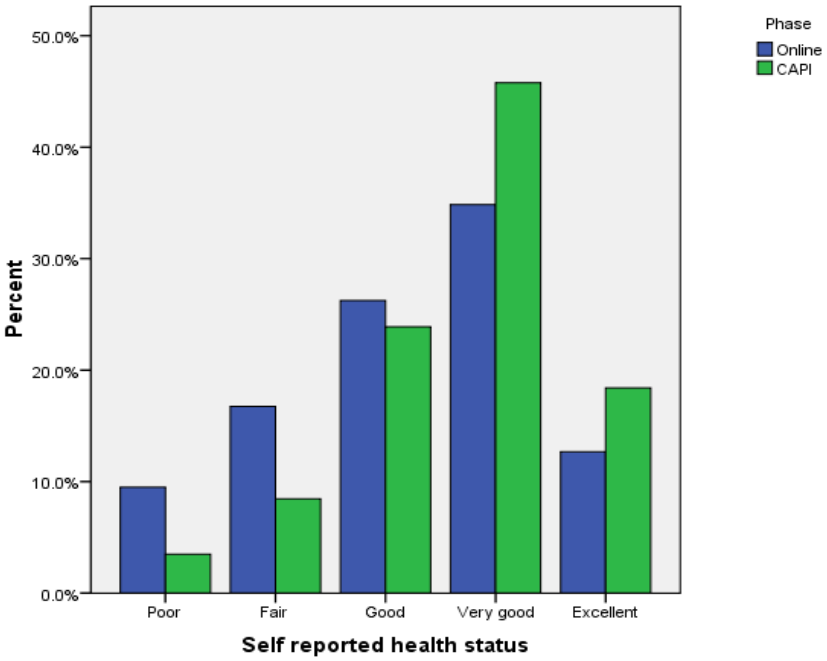


Figure 2: Self reported health satisfaction

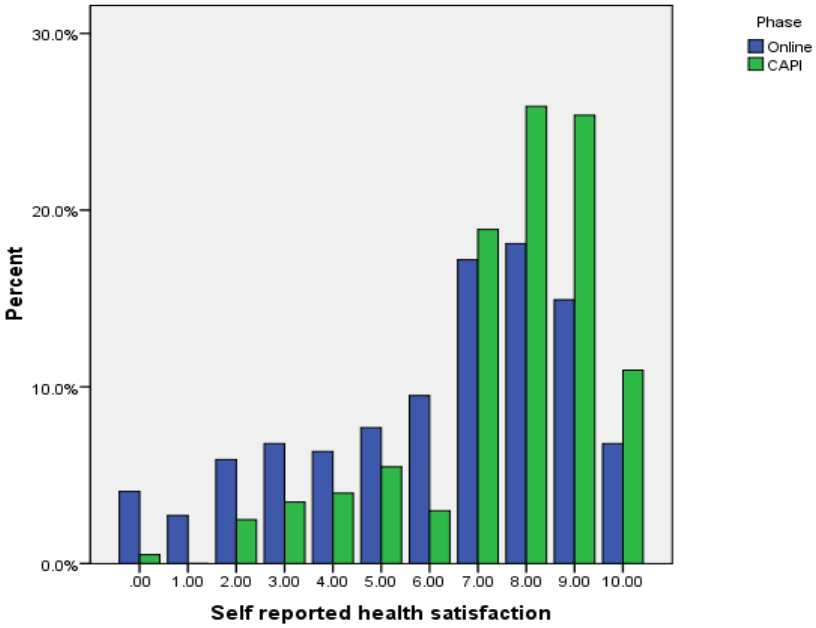


Figure 3: Self reported life satisfaction