EXPLORING THE SOCIAL VALUE OF HEALTH-CARE INTERVENTIONS: A STATED PREFERENCE DISCRETE CHOICE EXPERIMENT

COLIN GREENa,* and KAREN GERARDb

aInstitute of Health Service Research, Peninsula Medical School, Universities of Exeter and Plymouth, UK
bSchool of Nursing and Midwifery, University of Southampton, UK

SUMMARY

Much of the literature on distributive preferences covers specific considerations in isolation, and recent reviews have suggested that research is required to inform on the relative importance of various key considerations. Responding to this research recommendation, we explore the distributive preferences of the general public using a set of generic social value judgments. We report on a discrete choice experiment (DCE) survey, using face-to-face interviews, in a sample of the general population (n = 259). The context for the survey was resource allocation decisions in the UK National Health Service, using the process of health technology appraisal as an example. The attributes used covered health improvement, value for money, severity of health, and availability of other treatments, and it is the first such survey to use cost-effectiveness in scenarios described to the general public. Results support the feasibility and acceptability of the DCE approach for the elicitation of public preferences. Choice data are used to consider the relative importance of changes across attribute levels, and to model utility scores and relative probabilities for the full set of combinations of attributes and levels in the experimental design used (n = 64). Results allow the relative social value of health technology scenarios to be explored. Findings add to a sparse literature on ‘social’ preferences, and show that DCE data can be used to consider the strength of preference over alternative scenarios in a priority-setting context. Copyright © 2008 John Wiley & Sons, Ltd.

Received 15 November 2007; Revised 20 May 2008; Accepted 14 August 2008

KEY WORDS: priority setting; discrete choice experiments; public preference; methodology

1. INTRODUCTION

In health-care systems around the world decision makers are faced with competing demands and insufficient resources, even in the richest of countries (Hauck et al., 2003). In these circumstances it is not possible to provide all available and potentially beneficial health care to those who could benefit from it, and priority setting is inevitable. This is the case in the UK National Health Service (NHS), where in some cases access to effective health care is restricted or denied (New, 1997; Martin et al., 2002; Ham and Robert, 2003; Newdick, 2005).

In the UK, health policy makers at local or regional levels, or national bodies such as the National Institute for Health and Clinical Excellence (NICE), are responsible for making difficult resource allocation decisions. The basis for these decisions is often not clear (Ham, 1997; Hope, 2001; BMA, 2001; Birch and Gafni, 2002; Newdick, 2005; Dakin et al., 2006). One of the policy objectives of the UK...
NHS, when setting such priorities, is to take into account the views of the general public (Department of Health, 1997, 2000, 2001). Such a policy objective is evidenced by NICE, which states that ‘Underlying all [NICE] decisions...is one fundamental social value judgment: that advice from NICE to the NHS should embody values that are generally held by the population of the NHS’ (Rawlins and Culyer, 2004, p. 226).

While public involvement in health-care decision making is a policy objective in the UK, there is an absence of empirical evidence on how the public may value different health-care interventions (Sassi et al., 2001a,b; Schwappach, 2002a; Dolan et al., 2005). Recent reviews of the health-related literature around social values and distributive preferences have recommended that research is needed into the relationship (i.e. relative values) between key social values (Sassi et al., 2001a; Schwappach, 2002a; Hauck et al., 2003; Dolan et al., 2005).

This paper presents findings from a study to elicit the preferences of the general public over scenarios describing health-care interventions against generic social value judgments. The study uses the discrete choice experiment (DCE) framework to consider public preferences, and it specifically explores how the general public may weigh up competing distributive preferences in a priority-setting context. The study sets out to test the use of the DCE framework when used with samples of the general public in a priority-setting context, and an important aspect of the study is to present results in a way that is helpful in a health policy context.

### 2. USING DCEs TO EXPLORE SOCIAL VALUES

DCEs are a stated preference technique. They elicit people’s preferences, based on their stated preferences in hypothetical choices (Louviere et al., 2000). In the choice sets, people are presented with competing alternative scenarios (e.g. goods or services), described using defined attributes and a range of attribute levels, and they are asked to make a choice between them.

The DCE approach can provide a rich and versatile source of stated preference information. It is able to generate information to indicate whether particular attributes (e.g. specific social values) are a predictor of choice, over alternative scenarios, and it is able to provide information on the relative importance of the attributes used to describe the alternatives in choice sets (Louviere et al., 2000). The DCE is also able to generate information to indicate the relative overall importance of specific scenarios (combinations of attributes and levels) that are regarded as plausible competing scenarios (Louviere et al., 2000).

DCEs have been used widely to elicit preferences in a number of areas (Louviere et al., 2000), and they have been increasingly used in applied health economics research (Ryan et al., 2008). Ryan and Gerard (2003) and Fiebig et al., (2005) have presented useful reviews of the DCE studies published up to 2005. There is a debate over the advantages and disadvantages of the DCE approach, as applied in health care (e.g. Lloyd, 2003; Ryan and Amaya-Amaya, 2004; Bryan and Dolan, 2004). However, the feasibility and acceptability of the DCE approach is encouraging (Ryan et al., 2001; Ryan and Gerard, 2003; Hensher et al., 2005), and one of the advantages of the DCE approach is that it is able to present choices that are relevant to the respondent (Louviere et al., 2000).

Discrete choice studies have predominantly elicited individual values from respondents who have been asked to consider the choice context as it applies to themselves. There have been a small number of studies that have explored preferences in a ‘social’ context, where respondents have been asked for their preferences when considering choices (i.e. resource allocation and priority-setting choices) involving other people in society (e.g. Roberts et al., 1999; Farrar et al., 2000; Ratcliffe, 2000; Bryan et al., 2002; Schwappach, 2003; Gyrd-Hansen, 2004; Schwappach and Strasmann, 2006; Baltussen et al., 2006).

These ‘social’ DCE studies have been reviewed elsewhere (Green, 2007); see a summary of the characteristics of these studies in Appendix A. Across the 11 social context DCEs identified in the...
review by Green (2007), there are a range of study objectives explored, and a broad range of attributes and methods employed. In 4 of the 11 studies the context was one involving general priority-setting criteria, for policy-level considerations (Schwappach, 2003; Schwappach and Strasmann, 2006; Baltussen et al., 2006; Tappenden et al., 2007), as in the study presented in this paper. All four of these studies used self-complete methods, in selected or convenience samples. Other studies do address priority-setting dilemmas, but in those cases the context is more specific. For example, Ratcliffe (2000) considers distribution of donor organs, and Farrar et al., (2000) consider a hospital service development perspective.

Four of the 11 studies used interview methods, two in the UK (Roberts et al., 1999; Bryan et al., 2002) and two in Denmark (Gyrd-Hansen and Slothuus, 2002; Gyrd-Hansen, 2004). The UK studies were methodological in nature, investigating the assumptions underlying the maximisation of quality-adjusted life-years (QALYs), and comprised a small pilot study (n = 91) and a large random sample of the general population (n = 909). Two studies have considered cost-effectiveness (Baltussen et al., 2006; Tappenden et al., 2007), with both of these studies using a self-complete format in samples of health policy makers with some familiarity with the use of cost-effectiveness analysis.

The literature on social DCE studies is growing, and a number of different approaches have been taken. Yet, none of the reported studies have derived an index of social preference, from the statistical modelling techniques applied in discrete choice analysis, to reflect the distributive preferences of the general public, this being one of the outputs from the study presented here.

3. DCE FRAMEWORK

In general, the DCE approach involves a blend of economic theory, econometric methods, and survey design theory (psychology and statistical design). The approach is broadly consistent with microeconomic theory of consumer behaviour, and it draws upon Lancaster’s theory of consumer choice (Lancaster, 1966, 1971) and random utility theory (RUT) (McFadden, 1973; Hanemann, 1984). In discrete choice modeling, the utility function is composed of a deterministic component (V) and an unobservable or stochastic (random) component (ε):

$$U_{in} = V_{in} + \varepsilon_{in}$$  \hspace{1cm} (1)

where $V_{in}$ is the deterministic component of the utility for individual $n$ and option $i$ and $\varepsilon_{in}$ is the random or unobservable component for individual $n$ and option $i$. $V_{in}$ is the indirect utility function in which the attributes are arguments. Assuming a linear utility function (as is most commonly done), $V$ is therefore characterised as

$$V_{in} = \sum_i \beta_i X_{in} + \sum_p \psi_p W_{pn}$$  \hspace{1cm} (2)

where $V$ is the utility, $X$ is a matrix of attribute levels, $W$ is a vector of $p$ individual characteristics, $\beta_i$ are the coefficient estimates for each attribute in the matrix $X$ (marginal utilities), and the $\psi_p$ coefficients represent the extent to which personal characteristics influence choice.

Consistent with RUT (McFadden, 1973), the decision-making process within the DCE framework is seen as involving a comparison between indirect utility functions. It is assumed that respondents will consider all the relevant information and select the alternative with the highest utility. This allows the estimation of utility functions, and the estimation of the probability that alternatives will be chosen, from the broader choice set.

Given the underlying theoretical foundations, and the empirical evidence to support the use of DCEs for the elicitation of preferences over different attributes (Ryan and Gerard, 2003; Fiebig et al., 2005),
the DCE framework was regarded as a promising and appropriate approach to consider the relationship between key social value judgments, relevant in a priority-setting context. Furthermore, the modelling of discrete choice data, using the conditional logistic regression model, was seen as a means of providing a summary measure of the overall relative social value, against specific health-care intervention scenarios, using the link it provides between probabilistic and choice theory (McFadden, 1973). The approach allows the use of a probabilistic logistic model to consider the differences between utility functions, linking the probabilistic model to the model of choice.

Assuming that the \( \varepsilon \) term is independently and identically distributed (i.e. using a Gumbel distribution), the logistic distribution and the conditional logit model can be used to derive probability outcomes across a choice set (Louviere et al., 2000; Hensher et al., 2005). Characterising the probability of the \( i \)th scenario being chosen from the complete choice set \( \{j = 1, \ldots, n\} \), the conditional logit model takes the form:

\[
\Pr(i) = \frac{e^{V_i}}{\sum_j e^{V_j}} \quad (C \text{ of } j = 1, \ldots, n)
\]

4. METHODS

A general population DCE survey was conducted, using face-to-face interviews, in order to explore the preferences of the general public over health-care scenarios.

4.1. Setting

Health technology appraisal was used in the study as an example of a resource allocation problem in the UK NHS. Health technology appraisal is a common process (Oliver et al., 2004). It is used to appraise new and existing health technologies (e.g. drugs, devices, surgical procedures, different approaches to service delivery) in order to provide guidance on the overall value of these technologies to a particular population, subject to the objectives of the organisation or the health-care system it is servicing. In the UK, health-care resources must be assigned to competing demands, and decision makers have to decide whether health technologies can be recommended as a cost-effective use of NHS resources (e.g. National Institute for Clinical Excellence, 2004). Therefore, the process of health technology appraisal is undertaken either explicitly or implicitly, at all levels of the NHS to inform judgments on the coverage of health systems. Technology appraisal is regarded in the current study as a useful context to explore social values, drawing on the literature around the NICE health technology appraisal process, and respondents were asked to make a series of choices between descriptions of alternative health technology scenarios. Although the priority-setting context for health care is not an everyday setting for members of the general public, the DCE studies by Roberts et al., (1999), and Bryan et al. (2002), have demonstrated a good level of acceptability in a large sample of the UK general public. Therefore, in the current study, respondents were expected to be able to consider the nature of the choices presented, and to have preferences over the different scenarios presented.

4.2. Attributes and levels

The selection of the attributes was informed by: an extensive review of the empirical literature on distributive preferences (Green, 2007); a review of policy documents on the NICE health technology appraisal process (e.g. National Institute for Clinical Excellence, 2004) and published NICE guidance (www.nice.org); expert opinion through discussion with methods experts and health-care decision makers; and pilot work to test the best way to present information.
Much of the empirical literature is context specific (e.g. organ transplant decisions), and while it is important to consider context, in specific situations, a key objective (challenge) for the current study was to examine the general relationship between key social values and the public preferences for the allocation of resources across a wide range of health technologies. This was done using generic attributes and a non-labelled experimental design.

Table I describes the attributes and levels used. The four attributes used in the survey are all able to reflect generic social value judgments. That is, they are all indicating, and could be presented as a normative proposition, that some particular action ought to be implemented and is itself regarded as 'good' (good for society) in the context of resource allocation decisions e.g. it is good to use resources efficiently. These four attributes were judged to be the most appropriate generic attributes (i.e. not context specific) based on the literature review, the technology appraisal context being used, and the evidence from pre-pilot and formal pilot research.

The findings from a formal pilot study, including a full interview schedule in 25 respondents, supported the acceptability of the text and terms used in the survey and the feasibility of the proposed methods (e.g. choice scenarios presented, showcards used). In the pilot study all respondents stated that they were able to understand and complete questions, and interviewer feedback was positive against level of understanding and respondent engagement.

In selecting the attributes, the literature was reviewed in detail, including experimental studies, attitudinal studies, and the subsets of literature, which presented detailed interview based, or focus group based, qualitative research (e.g. Dicker and Armstrong, 1995; Stronks et al., 1997; Cookson and Dolan, 1999). It was clear from the literature reviewed, and from previously published reviews (Sassi et al., 2001a; Schwappach, 2002a; Dolan et al., 2005), that some commonly cited considerations, e.g. age, desert/lifestyle, did not have strong empirical support as priority-setting criteria. For example, on age (probably one of the most contentious exclusions from the design here), the empirical evidence is inconclusive, and the policy context provided a clear guide that age was not supported as a basis for setting priorities (National Institute for Health and Clinical Excellence, 2005). Reviews of the empirical evidence have concluded that evidence on the importance of age in priority setting may be interpreted as being supportive of giving lesser weight to older people, versus younger people, but that the empirical evidence is simple, often using crude methods, and that findings may be related to other confounding factors (e.g. capacity to benefit, larger health gains) and may not be a direct reflection of a preference against a specific age group (Sassi et al., 2001a,b; Schwappach, 2002a; Dolan et al., 2005). At a national policy level in the UK, NICE does not support the use of age as a basis for setting priorities in its technology appraisal process (National Institute for Health and Clinical Excellence, 2005; Rawlins and

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Description (summary)</th>
<th>Levels</th>
</tr>
</thead>
<tbody>
<tr>
<td>Severity</td>
<td>Whether patients are severely affected by their condition</td>
<td>Yes, No</td>
</tr>
<tr>
<td>Health improvement</td>
<td>The average health improvement expected from the treatment</td>
<td>Large, Moderate, Small, Very small</td>
</tr>
<tr>
<td>Value for money</td>
<td>Cost-effectiveness of treatment – the value for money expected from the treatment</td>
<td>Very good, Fairly good, Fairly poor, Very poor</td>
</tr>
<tr>
<td>Other treatment</td>
<td>Whether other effective treatments are available for the patient group</td>
<td>Yes, No</td>
</tr>
</tbody>
</table>

See ‘showcard’ used in the survey in Appendix B. The ‘least desirable’ scenario is expected to be where Severity = no, Health improvement = very small, Value for money = very poor, Other treatment = yes.
Culyer, 2004). On age, and on other issues, such as health-related behaviour (desert/lifestyle), studies can be selected in isolation to support specific viewpoints, but on balance there is no clear indication of the extent to which these factors should be used in a priority-setting context.

The broad process of health technology appraisal is only relevant when a technology has been shown to be effective (Culyer, 2001), and to offer benefits, but a measure of effectiveness (with levels for effectiveness, e.g. very effective, modest effectiveness) was not considered as a relevant attribute in the study design. Judgments on differing levels of effectiveness are often dependent on condition-specific outcomes (e.g. progression-free survival, impact on biomarkers such as blood glucose or lung function), and this hampers comparisons in a generic context. Given the health technology appraisal context used in the study, the importance attached to the level of effectiveness and its impact on the health of those treated was expected to be captured in the attribute for level of health improvement. Again based on the setting used, issues surrounding the number of people treated and impact on the health-care budget (affordability) were not considered for inclusion in the study design, as these are not relevant considerations in the NICE technology appraisal process (National Institute for Health and Clinical Excellence, 2005).

Attributes used covered severity of health, level of health improvement, value for money, and the availability of other treatments. A primary consideration when designing the study was the inclusion of an attribute to capture the level of cost-effectiveness. Cost-effectiveness is a prominent consideration in priority-setting problems in the UK, and none of the published ‘social’ DCE surveys, in samples of the general population, have used cost-effectiveness as an attribute in the survey design. Therefore, a specific objective here was to test the DCE framework, with cost-effectiveness directly presented in the design, in a survey of the general public. As the study was designed to test this issue, and given a strong a priori emphasis on the translation of the DCE results in a policy-relevant way, the experimental design chosen was quite simple, with only the four attributes (across either two or four levels). In 8 of the 11 social DCEs identified in the literature reviewed, between 4 and 6 attributes have been used. The study by Bryan et al. (2002), which showed the DCE approach to be feasible and acceptable in a sample of the UK population, used four attributes (with between two and four levels). This current study is an exploratory study, testing the application of a DCE in a challenging and important social context. However, the attributes and levels used were still expected to provide empirical evidence that would be meaningful in a health policy setting.

4.2.1. Severity. Severity of health, the pre-treatment health state of patients, is identified in the current literature as a social value that is supported by respondents in a number of empirical studies reporting experimental data (e.g. Nord, 1993a,b; Ubel, 1999; Edwards et al., 2003; Gyrd-Hansen, 2004; Nord et al., 2005) and in studies reporting attitudinal data (e.g. Bowling et al., 2002; Myllykangas et al., 2003). In the policy literature, severity of health condition is not specifically referred to by NICE, in its report on social values (National Institute for Health and Clinical Excellence, 2005), but in the commentary literature around NICE decision making it is indicated that severity of health condition may be a consideration in the decision-making process (e.g. Devlin and Parkin, 2004).

There are a wide variety of ways to describe severity of health. In this study it was important to keep the meaning generalisable across health states. Therefore, the descriptive system of the generic quality of life instrument, the EuroQol (EQ-5D) (EuroQol Group, 1990), was used. The DCE study used text from EQ-5D health state descriptions to explain the notion of severity (see Appendix A). The attribute was assigned two levels (yes/no), and a priori it was expected that health interventions for those patient groups who were more severely affected by their health condition (all else equal) would be associated with a greater level of utility.

4.2.2. Value for money. In the health policy literature, efficiency is a well-supported motive in the allocation of health resources (e.g. Sassi et al., 2001b), and the use of cost-effectiveness analysis in health
care is now widespread. The UK Department of Health has to ensure that public money is spent wisely and efficiently, with the expenditure on health and social care expected to represent value for money (www.dh.gov.uk). It follows that the NHS places the cost-effectiveness of treatment as a prime consideration for both the assessment and appraisal of interventions respectively (e.g. National Institute for Clinical Excellence, 2004).

Cost-effectiveness is a stated objective in the NICE technology appraisal process (National Institute for Clinical Excellence, 2004), and NICE has offered some guidance on cost-effectiveness, in terms of the cost per QALY (National Institute for Health and Clinical Excellence, 2005). However, the terminology of cost-effectiveness and the efficiency concept of the cost per QALY are not commonplace for the general public, with respect to health care. Therefore, the term ‘value for money’ was used to express the notion of cost-effectiveness and efficiency (Appendix B). This term was regarded as a commonly understood term and very much related to efficiency and cost-effectiveness. Public health decision makers indicated that the terminology of value for money had been used successfully when dealing with the public. The attribute was assigned four levels, and it was expected a priori that health interventions with better levels of value for money (all else equal) would be associated with a greater level of utility.

4.2.3. Improvement in health. Health improvement is a key health policy objective. The evidence base reviewed demonstrated that health gain, and the level of health gain, is an important consideration for respondents when eliciting preferences for the allocation of health-care resources (e.g. Charny et al., 1989; Bowling, 1996; Cookson and Dolan, 1999; Edwards et al., 2003). While this attribute is linked with the attribute of ‘value for money’, it is clear from the empirical literature reviewed, and from the literature around the cost-effectiveness of health technologies, that it is possible to have different levels of health improvement, at any of the levels used for value for money. For example, a large health improvement (or a very small health improvement) can be either very good or very poor value for money, supporting the treatment of the attributes as separate social value judgments. While acknowledging the potential for attributes to be related, there is no unambiguous empirical evidence to know whether and to what extent such interactions may be present.

The attribute was assigned four levels, and it was expected a priori that health interventions with larger health improvements (all else equal) would be associated with a greater level of utility.

4.2.4. Other treatments. When considering the allocation of resources to a health technology being appraised, this attribute establishes whether a patient group has any other effective treatment options available (other than the option being appraised) or whether the current standard treatment is best supportive care. The attribute reflects the importance of general circumstances surrounding the availability or otherwise of alternative treatments. The empirical literature does not inform on this attribute in any detail, but there is support for the use of the attribute from the health policy literature. Studies by Williams et al. (2005) and Tappenden et al. (2007) have suggested the potential value of this attribute, in a priority-setting context. Also, when reviewing the published NICE technology appraisal guidance, the availability of other treatments, or not, was information frequently presented in the final reports of the appraisal considerations. For example, with reference to the appraisal of imatinab for the treatment of chronic myeloid leukaemia (National Institute for Clinical Excellence, 2003), Rawlins and Culyer (2004) state that although the cost-effectiveness of treatment in this case was outside of the range usually acceptable to NICE, ‘in the absence of any effective alternative treatment…imatinab was considered to be cost-effective in the chronic phase after interferon alfa’ (p. 225, italics not in original). The attribute was assigned two levels, and it was expected a priori that health interventions introduced when there were no other effective treatments available (all else equal) would be associated with a greater level of utility.
4.3. Experimental design

The experimental design used has a full factorial of 64 combinations of attributes and levels ($4^2 \times 2^2$). Pairwise choice sets were constructed using contemporary experimental design theory (Street et al., 2008). In DCE design there is now greater awareness of alternative classes of designs with known statistical efficiency, usually measured as D-efficiency or D-error to indicate level of inefficiency based on the asymptotic variance–covariance matrix of the parameter estimates. The study here followed the approach recommended by Street and Burgess (2003, 2004). This takes a fractional orthogonal main effects plan with known efficiency and uses fold-over copies to create the necessary subsequent choices in the choice set. In this way a sample of 16 pairwise choices, giving a choice set with 94% D-efficiency, were obtained. This design assumes that interactions among attributes are insignificant in all two-way and higher-order interactions. This is considered to be a reasonable assumption, as 70–90% of variance can be explained by main effects attributes only (Dawes and Corrigan, 1974; Louviere et al., 2000). The design used had level balance (Huber and Zwerina, 1996) and was orthogonal, minimising multicollinearity.

All of the scenarios used were checked for plausibility i.e. the potential that health technologies would fit the scenarios. To limit respondent burden, choices were blocked into two sets of eight choices, for two versions of the interview questionnaire. The blocking into two sets of questions was undertaken using an additional attribute column from the factorial design employed, to ensure orthogonality over the choice set.

4.4. Scenario presentation

Respondents were asked to make a series of forced choices involving two alternative health-care scenarios. Scenarios (attributes and levels) were presented as characteristics of health technologies. Respondents were asked to put themselves in the position (context) of an NHS decision maker faced with difficult priority-setting decisions. They were informed that these difficult decisions had to be made, and one of the two alternative scenarios (in each choice set) had to be chosen, as the decision maker was unable to fund all of the health technologies it would wish to (see Figure 1, interviewer text, for example).

Question: SHOWCARD A health care decision maker is faced with difficult choices on how to allocate its budget. Imagine a choice where there are two options for the use of available funds. Given that only one of the options can receive funding, which option would you support?

Option K Patients are severely affected by their condition. With treatment the average patient has a very small improvement in their health. The treatment is regarded as being very poor value for money. There are no other effective treatment options available.

Option L Patients are not severely affected by their condition. With treatment the average patient has a large improvement in their health. The treatment is regarded as being very good value for money. There are other effective treatment options available.

<table>
<thead>
<tr>
<th>Option</th>
<th>Response</th>
</tr>
</thead>
<tbody>
<tr>
<td>Option K</td>
<td>1</td>
</tr>
<tr>
<td>Option L</td>
<td>2</td>
</tr>
<tr>
<td>Other</td>
<td>3</td>
</tr>
<tr>
<td>Don’t Know/Unable to choose</td>
<td>4</td>
</tr>
<tr>
<td>None</td>
<td>5</td>
</tr>
<tr>
<td>Refused</td>
<td>6</td>
</tr>
</tbody>
</table>

Figure 1. Interviewer text (an example of DCE question)
Given that the aim of the study was to explore preferences over generic attributes and health-care scenarios, descriptive detail on disease, condition, or type of health technology was not used in the DCE design. Pilot study results indicated that this approach was feasible and acceptable to respondents (Green, 2007). Findings from earlier DCE studies in samples of the general public supported a generic presentation of health-care scenarios (Roberts et al., 1999; Bryan et al., 2002).

Interviews involved background and introductory information, including a description of the attributes (see Appendix B for showcard). To assess level of understanding with the DCE format, a ‘consistency check’ question was used, where one of the two options in the choice set was regarded as a dominant option (i.e. all attribute levels were regarded as more desirable). Throughout the interview, interviewers talked through each of the DCE questions, and each question had a showcard to summarise the choice (see example in Figure 2).

**4.5. Sample/data collection**

A general population sample, from the Southampton (UK) City Council area, was used. A random-location quota sampling approach was employed. This sets fixed quotas of people to be interviewed in a number of randomly selected sampling points. Sampling points were based on ‘output areas’ (OAs), the smallest building block of the UK Census. OAs were randomly selected proportionate to population size, controlling for socio-demographic composition. Quotas were set – individually at each sample point – to reflect the socio-demographic profile of residents, on gender, age, and work status, using profile data from the 2001 Census. Face-to-face interviews were conducted by a public survey company (Ipsos MORI), using experienced interviewers (over September–October 2005). There is limited guidance on sample size calculations for discrete choice studies, and there are no practical well-designed rules to guide the analyst (Hensher et al., 2005). Pearmain et al., (1991) have suggested that for DCE designs sample sizes over 100 are able to provide a basis for modelling preference data, and Hensher...
et al. (2005) have suggested a rule of thumb of 50 respondents per question to provide adequate variation in the variables of interest. The sample size used here was based on resource constraints, with a sample of 250 respondents sought (and judged to be credible), and importance was placed on getting a sample that was representative of the general population. Each interview had four sections:

1. Introduction/rationale for survey, ranking of attributes (for familiarity and warm-up).
2. DCE priority-setting questions (example, consistency question, and eight paired choices).
3. Questions on difficulty with understanding and answering DCE questions.
4. Socio-demographic questions (e.g. age, sex, employment status, household characteristics) and questions on current health status, experience of illness.

Interviewers (Ipsos MORI) assigned one of the two versions of the survey instrument using a rotation random assignment method, in each area, when undertaking interviews.

4.6. Data analysis

Choice data were modelled using a random utility maximisation framework (Louviere et al., 2000; Hensher et al., 2005) and STATA 8.1 software (STATA Corporation, 2003). As the data are binary choice data – ‘1’ represents the option being chosen, with ‘0’ where not chosen – the conditional logit model is used for modelling. A series of utility models were fitted (main effects, main effects with interactions). Logit models were considered the most appropriate given the decision-making context. That is, (i) logit regression coefficients reflect a log of the odds and transform directly into an odds ratio, to directly reflect the impact and relative effect of each attribute (level), (ii) logit models use the ‘S-shaped’ logistic probability distribution (Greene, 2003), which may be regarded as the more appropriate approach for the estimation of probabilities (central to the current analysis).

The model estimated is of the form

\[ V = \beta_0 + \beta_1 \text{Severity} + \beta_2 \text{Other treatments} + \beta_{3-5} \text{Health improvement} + \beta_{6-8} \text{Value for money} \]

This is a linear in parameters main effects utility function, which is the functional form that has been used in over two-thirds of the DCE studies in the health economics literature (Fiebig et al., 2005). The deterministic component of the utility function \( V \) is a function of the attribute levels between options, where the coefficients (part-worth utilities) \( \beta_1-\beta_8 \) and the constant \( \beta_0 \) are estimated in the model. The part-worth utilities \( \beta \) can be summed to give an overall utility for each combination of attribute levels. This gives an indication of the relative social value of the scenarios in the experimental design, and allows consideration of the impact of changes in single attribute levels and combinations of attribute-level changes on the health technology scenarios described compared with a reference case.

Attributes are coded so that the constant term \( \beta_0 \) is used to reflect what is expected to be the least desirable option (for funding) in the factorial design (see footnote to Table I). Dummy variables are used to account for this approach, and the least desirable option defined acts as a reference case. A priori the coefficients \( \beta_1-\beta_8 \) are expected to have a positive sign, indicating an increase in utility (probability of being chosen for funding) relative to the reference case (constant). The use of dummy variables is helpful in interpreting the findings of the study, at a policy level, providing a meaningful reference case for a health policy audience.

Modelling considered subgroups (i.e. interaction effects against respondent characteristics) defined against (i) the age of respondents, (ii) their self-reported health status, and (iii) their experience of illness. These subgroups were defined a priori based on a review of the DCE literature (Green, 2007).
Data from the logit model were transformed into a set of probabilities, to show the relative probabilities of being chosen/supported by the general public (based on the sample used) for each of the 64 scenarios of the full factorial experimental design. This approach is consistent with the conditional logistic regression approach discussed above (Equation (3)).

5. RESULTS

A sample of 263 respondents were interviewed. Four respondents did not provide DCE response data for any of the eight DCE questions, indicating ‘don’t know’ (or with response not stated) on the questionnaire. These four respondents are excluded from analyses. The usable sample included 259 respondents, giving a potential data set of 2072 pairwise observations \( n = 259 \times 8 \) questions. A small number of respondents \( n = 26 \) did not provide a preference between the two options presented in each of the 8 questions, and the resulting data set was 2027 observations (98%).

This sample provided a representative sample of the socio-demographic characteristics of the Southampton City Council geographic area (Southampton City Council, Southampton 2004) (Table II). However, the sample did have a slightly larger proportion of retired and home workers, than in the population at large, which may be expected in an ‘in-home’ interview survey of this type. There were no statistically significant differences in socio-demographic characteristics across the two versions of the questionnaire used \( \chi^2 \) tests).

Twelve respondents ‘failed’ a preliminary ‘consistency check’ question, choosing the option that was dominated across all attributes by the alternative available. In two further cases the response data for this question stated ‘don’t know’ or ‘not-stated’. Therefore, in 5% of cases (14/259) the sample did not answer the consistency question as expected. Given the societal context of the survey, and a lack of clarity in the research literature on what might be classed as irrational response data (Lanscar and Louviere, 2006), the data from these respondents have been included in the analysis.

Table III reports the results of the conditional logit model, for analysis of main effects (Model A) and main effects with interactions for respondent characteristics (Model B). This latter analysis has been used in this paper for the presentation and discussion of results. The model demonstrates a good fit, in terms of the pseudo-\( R^2 \) statistic, and predictive power. All main effects coefficients are positive (as per \textit{a priori} expectations) and the staging of the impact of attributes, across attribute levels, is consistent (as expected) over the incremental changes in each attribute (e.g. \( \beta_3 < \beta_4 < \beta_5 \)). In Model B, which includes an interaction effect, introduced for self-reported health status, all of the main effects are statistically significant. This indicates that the choices respondents made were sensitive to variation in the levels for the attributes. The findings provide support for the theoretical validity of the model. The interaction effect, against self-reported health status (and the ‘other treatment’ attribute), is introduced into the model (estimation of the utility function) by applying the proportion of persons in the sample (9.3%) with self-reported ‘bad’ health status, which reduces the impact of the coefficient stated in Table III when estimating a utility function (i.e. coefficient is multiplied by proportion). When respondent characteristics covering age categories and inputs for experience of illness were introduced into the model, there was no indication from model output (likelihood ratio tests) that the variables have a significant impact on the fit of the model.

The coefficients in Table III reflect the part-worth utilities associated with changes in each of the attribute levels (compared with the reference case). For example, a health technology scenario with a large health improvement has a greater level of utility (social value) than scenarios with a small or a moderate health improvement, all else equal. Table III also presents the odds ratios (odds of success, i.e. of being chosen/funded) against attribute levels to provide a further interpretation of the relative importance of incremental changes in attribute levels from the reference case scenario.
Table IV presents utility (preference) scores for all combinations of attribute levels in the full factorial design (64 scenarios), with scores ranging from a high of 3.44 to a low of 0.23. The differences in utility between scenarios demonstrate the impact of attribute-level changes on the relative social value for each health technology scenario, showing the most desirable combinations of attribute levels.

Table IV also presents an estimate of the ‘probability’ of success (i.e. being selected for funding) for each of the scenarios, relative to one another (i.e. probability of being selected from the total group of 64 options). The probabilities for the 64 scenarios sum to 1.00, with the top-ranked item showing a probability of 5.39% of being chosen from the set of 64 scenarios and the bottom-ranked item having a probability of 0.22%. These probabilities provide a measure of preference, with higher probabilities indicating a greater likelihood of selection.
for funding the described scenario) showing that the health technology scenarios are given higher priority (preference) compared with those with a lower probability of being chosen for funding (from the full set of scenarios available).

Table V presents data on respondent self-reported difficulty in understanding and answering the DCE questions posed. The majority of respondents reported that they found the questions not very or not at all difficult to understand. However, the majority of respondents did report that the questions were very or fairly difficult to answer.

There were no statistically significant differences ($\chi^2$ test) in responses to difficulty questions by questionnaire version (1 and 2). Generally there were no significant differences in response to difficulty questions by respondent characteristics. There were some significant differences in understanding the questions by age, with the elderly likely to find questions more difficult to understand, and differences by self-reported health status. For health status groups (self-reported), those in bad or very bad health were more likely (16 of 24) to have difficulty in understanding the questions ($p$-value $= 0.011$). There were no statistically significant differences (nor a tendency towards a difference) between subgroups in terms of difficulty in answering the questions.

6. DISCUSSION

When making priority-setting decisions policy makers are often faced with difficult decisions between options, which are all regarded as potentially beneficial. However, there are often a range of social values that influence policy decisions, and when there are trade-offs across these social values, or equity arguments (objectives), judgments are needed on which things are better than others (referred to as ‘betterness’ by Broome, 1991). Reviews in the health-care literature have identified the need to investigate the relationship between key equity arguments (Sassi et al., 2001a; Schwappach, 2002a; Dolan et al., 2005). The findings presented in this paper provide a number of opportunities to weigh different social values against one another, in a policy-relevant context, and to consider which health technology scenarios are better than others, from the perspective of the social preferences elicited.

At an attribute (or social value judgment) level, the results show that the most important changes in attribute levels in the choice model i.e. changes in utility for health technology scenarios, relative to the
Table IV. DCE model: estimated utility score and probability of success by scenario (full factorial)

<table>
<thead>
<tr>
<th>Ranked by probability</th>
<th>Value for money</th>
<th>Health improvement</th>
<th>Severity</th>
<th>Other treatments</th>
<th>Utility (logit)</th>
<th>Probability (%)</th>
<th>Cummulative probability (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Very good</td>
<td>Large</td>
<td>Severe</td>
<td>No other Tx</td>
<td>3.44</td>
<td>5.39</td>
<td>5.39</td>
</tr>
<tr>
<td>2</td>
<td>Very good</td>
<td>Large</td>
<td>Severe</td>
<td>Other Tx avail.</td>
<td>3.31</td>
<td>4.76</td>
<td>10.16</td>
</tr>
<tr>
<td>3</td>
<td>Fairly good</td>
<td>Large</td>
<td>Severe</td>
<td>No other Tx</td>
<td>3.30</td>
<td>4.68</td>
<td>14.83</td>
</tr>
<tr>
<td>4</td>
<td>Fairly good</td>
<td>Large</td>
<td>Severe</td>
<td>Other Tx avail.</td>
<td>3.17</td>
<td>4.13</td>
<td>18.96</td>
</tr>
<tr>
<td>5</td>
<td>Very good</td>
<td>Moderate</td>
<td>Severe</td>
<td>No other Tx</td>
<td>3.11</td>
<td>3.87</td>
<td>22.83</td>
</tr>
<tr>
<td>6</td>
<td>Very good</td>
<td>Moderate</td>
<td>Severe</td>
<td>Other Tx avail.</td>
<td>2.98</td>
<td>3.42</td>
<td>26.25</td>
</tr>
<tr>
<td>7</td>
<td>Fairly good</td>
<td>Moderate</td>
<td>Severe</td>
<td>No other Tx</td>
<td>2.96</td>
<td>3.36</td>
<td>29.61</td>
</tr>
<tr>
<td>8</td>
<td>Very good</td>
<td>Large</td>
<td>Not severe</td>
<td>No other Tx</td>
<td>2.91</td>
<td>3.17</td>
<td>32.78</td>
</tr>
<tr>
<td>9</td>
<td>Fairly good</td>
<td>Moderate</td>
<td>Severe</td>
<td>Other Tx avail.</td>
<td>2.84</td>
<td>2.96</td>
<td>35.75</td>
</tr>
<tr>
<td>10</td>
<td>Very good</td>
<td>Large</td>
<td>Not severe</td>
<td>Other Tx avail.</td>
<td>2.78</td>
<td>2.80</td>
<td>38.55</td>
</tr>
<tr>
<td>11</td>
<td>Fairly good</td>
<td>Large</td>
<td>Not severe</td>
<td>No other Tx</td>
<td>2.76</td>
<td>2.75</td>
<td>41.29</td>
</tr>
<tr>
<td>12</td>
<td>Fairly good</td>
<td>Large</td>
<td>Not severe</td>
<td>Other Tx avail.</td>
<td>2.64</td>
<td>2.43</td>
<td>43.72</td>
</tr>
<tr>
<td>13</td>
<td>Fairly poor</td>
<td>Large</td>
<td>Severe</td>
<td>No other Tx</td>
<td>2.58</td>
<td>2.28</td>
<td>46.00</td>
</tr>
<tr>
<td>14</td>
<td>Very good</td>
<td>Moderate</td>
<td>Not severe</td>
<td>No other Tx</td>
<td>2.58</td>
<td>2.28</td>
<td>48.28</td>
</tr>
<tr>
<td>15</td>
<td>Very good</td>
<td>Small</td>
<td>Severe</td>
<td>No other Tx</td>
<td>2.56</td>
<td>2.24</td>
<td>50.52</td>
</tr>
<tr>
<td>16</td>
<td>Fairly poor</td>
<td>Large</td>
<td>Severe</td>
<td>Other Tx avail.</td>
<td>2.46</td>
<td>2.02</td>
<td>52.54</td>
</tr>
<tr>
<td>17</td>
<td>Very good</td>
<td>Moderate</td>
<td>Not severe</td>
<td>Other Tx avail.</td>
<td>2.45</td>
<td>2.01</td>
<td>54.55</td>
</tr>
<tr>
<td>18</td>
<td>Very good</td>
<td>Small</td>
<td>Severe</td>
<td>Other Tx avail.</td>
<td>2.43</td>
<td>1.98</td>
<td>56.52</td>
</tr>
<tr>
<td>19</td>
<td>Fairly good</td>
<td>Moderate</td>
<td>Not severe</td>
<td>No other Tx</td>
<td>2.43</td>
<td>1.97</td>
<td>58.50</td>
</tr>
<tr>
<td>20</td>
<td>Fairly good</td>
<td>Small</td>
<td>Severe</td>
<td>No other Tx</td>
<td>2.42</td>
<td>1.94</td>
<td>60.43</td>
</tr>
<tr>
<td>21</td>
<td>Fairly good</td>
<td>Moderate</td>
<td>Not severe</td>
<td>Other Tx avail.</td>
<td>2.31</td>
<td>1.74</td>
<td>62.18</td>
</tr>
<tr>
<td>22</td>
<td>Fairly good</td>
<td>Small</td>
<td>Severe</td>
<td>Other Tx avail.</td>
<td>2.29</td>
<td>1.71</td>
<td>63.89</td>
</tr>
<tr>
<td>23</td>
<td>Very poor</td>
<td>Large</td>
<td>Severe</td>
<td>No other Tx</td>
<td>2.26</td>
<td>1.67</td>
<td>65.56</td>
</tr>
<tr>
<td>24</td>
<td>Fairly poor</td>
<td>Moderate</td>
<td>Severe</td>
<td>No other Tx</td>
<td>2.25</td>
<td>1.64</td>
<td>67.20</td>
</tr>
<tr>
<td>25</td>
<td>Very poor</td>
<td>Large</td>
<td>Severe</td>
<td>Other Tx avail.</td>
<td>2.14</td>
<td>1.47</td>
<td>68.67</td>
</tr>
<tr>
<td>26</td>
<td>Fairly poor</td>
<td>Moderate</td>
<td>Severe</td>
<td>Other Tx avail.</td>
<td>2.12</td>
<td>1.45</td>
<td>70.12</td>
</tr>
<tr>
<td>27</td>
<td>Very good</td>
<td>Very small</td>
<td>Severe</td>
<td>No other Tx</td>
<td>2.06</td>
<td>1.36</td>
<td>71.48</td>
</tr>
<tr>
<td>28</td>
<td>Fairly poor</td>
<td>Large</td>
<td>Not severe</td>
<td>No other Tx</td>
<td>2.05</td>
<td>1.34</td>
<td>72.82</td>
</tr>
<tr>
<td>29</td>
<td>Very good</td>
<td>Small</td>
<td>Not severe</td>
<td>No other Tx</td>
<td>2.03</td>
<td>1.32</td>
<td>74.14</td>
</tr>
<tr>
<td>30</td>
<td>Very good</td>
<td>Very small</td>
<td>Severe</td>
<td>Other Tx avail.</td>
<td>1.94</td>
<td>1.20</td>
<td>75.34</td>
</tr>
<tr>
<td>31</td>
<td>Very poor</td>
<td>Moderate</td>
<td>Severe</td>
<td>No other Tx</td>
<td>1.93</td>
<td>1.20</td>
<td>76.54</td>
</tr>
<tr>
<td>32</td>
<td>Fairly poor</td>
<td>Large</td>
<td>Not severe</td>
<td>Other Tx avail.</td>
<td>1.92</td>
<td>1.19</td>
<td>77.72</td>
</tr>
<tr>
<td>ID</td>
<td>Quality</td>
<td>Size</td>
<td>Severity</td>
<td>Other Tx Availability</td>
<td>V1</td>
<td>V2</td>
<td>V3</td>
</tr>
<tr>
<td>----</td>
<td>-----------</td>
<td>----------</td>
<td>-----------</td>
<td>-----------------------</td>
<td>-----</td>
<td>-----</td>
<td>-----</td>
</tr>
<tr>
<td>33</td>
<td>Fairly good</td>
<td>Very small</td>
<td>Severe</td>
<td>No other Tx</td>
<td>1.92</td>
<td>1.18</td>
<td>78.90</td>
</tr>
<tr>
<td>34</td>
<td>Very good</td>
<td>Small</td>
<td>Not severe</td>
<td>Other Tx avail.</td>
<td>1.90</td>
<td>1.16</td>
<td>80.07</td>
</tr>
<tr>
<td>35</td>
<td>Fairly good</td>
<td>Small</td>
<td>Not severe</td>
<td>No other Tx</td>
<td>1.88</td>
<td>1.14</td>
<td>81.21</td>
</tr>
<tr>
<td>36</td>
<td>Very poor</td>
<td>Moderate</td>
<td>Severe</td>
<td>Other Tx avail.</td>
<td>1.81</td>
<td>1.06</td>
<td>82.26</td>
</tr>
<tr>
<td>37</td>
<td>Fairly good</td>
<td>Very small</td>
<td>Severe</td>
<td>Other Tx avail.</td>
<td>1.80</td>
<td>1.04</td>
<td>83.31</td>
</tr>
<tr>
<td>38</td>
<td>Fairly good</td>
<td>Small</td>
<td>Not severe</td>
<td>Other Tx avail.</td>
<td>1.76</td>
<td>1.01</td>
<td>84.31</td>
</tr>
<tr>
<td>39</td>
<td>Very poor</td>
<td>Large</td>
<td>Not severe</td>
<td>No other Tx</td>
<td>1.73</td>
<td>0.98</td>
<td>85.29</td>
</tr>
<tr>
<td>40</td>
<td>Fairly poor</td>
<td>Moderate</td>
<td>Not severe</td>
<td>No other Tx</td>
<td>1.72</td>
<td>0.96</td>
<td>86.26</td>
</tr>
<tr>
<td>41</td>
<td>Fairly poor</td>
<td>Small</td>
<td>Severe</td>
<td>No other Tx</td>
<td>1.70</td>
<td>0.95</td>
<td>87.20</td>
</tr>
<tr>
<td>42</td>
<td>Very poor</td>
<td>Large</td>
<td>Not severe</td>
<td>Other Tx avail.</td>
<td>1.61</td>
<td>0.87</td>
<td>88.07</td>
</tr>
<tr>
<td>43</td>
<td>Fairly poor</td>
<td>Moderate</td>
<td>Severe</td>
<td>Other Tx avail.</td>
<td>1.59</td>
<td>0.85</td>
<td>88.92</td>
</tr>
<tr>
<td>44</td>
<td>Fairly poor</td>
<td>Small</td>
<td>Severe</td>
<td>Other Tx avail.</td>
<td>1.58</td>
<td>0.84</td>
<td>89.76</td>
</tr>
<tr>
<td>45</td>
<td>Very good</td>
<td>Very small</td>
<td>Not severe</td>
<td>No other Tx</td>
<td>1.53</td>
<td>0.80</td>
<td>90.56</td>
</tr>
<tr>
<td>46</td>
<td>Very good</td>
<td>Very small</td>
<td>Not severe</td>
<td>Other Tx avail.</td>
<td>1.41</td>
<td>0.71</td>
<td>91.27</td>
</tr>
<tr>
<td>47</td>
<td>Very poor</td>
<td>Moderate</td>
<td>Not severe</td>
<td>No other Tx</td>
<td>1.40</td>
<td>0.70</td>
<td>91.97</td>
</tr>
<tr>
<td>48</td>
<td>Fairly good</td>
<td>Very small</td>
<td>Not severe</td>
<td>No other Tx</td>
<td>1.39</td>
<td>0.69</td>
<td>92.66</td>
</tr>
<tr>
<td>49</td>
<td>Very poor</td>
<td>Small</td>
<td>Severe</td>
<td>No other Tx</td>
<td>1.38</td>
<td>0.69</td>
<td>93.36</td>
</tr>
<tr>
<td>50</td>
<td>Very poor</td>
<td>Moderate</td>
<td>Not severe</td>
<td>Other Tx avail.</td>
<td>1.28</td>
<td>0.62</td>
<td>93.98</td>
</tr>
<tr>
<td>51</td>
<td>Fairly good</td>
<td>Very small</td>
<td>Not severe</td>
<td>Other Tx avail.</td>
<td>1.26</td>
<td>0.61</td>
<td>94.59</td>
</tr>
<tr>
<td>52</td>
<td>Very poor</td>
<td>Small</td>
<td>Severe</td>
<td>Other Tx avail.</td>
<td>1.26</td>
<td>0.61</td>
<td>95.20</td>
</tr>
<tr>
<td>53</td>
<td>Fairly poor</td>
<td>Very small</td>
<td>Severe</td>
<td>No other Tx</td>
<td>1.20</td>
<td>0.58</td>
<td>95.78</td>
</tr>
<tr>
<td>54</td>
<td>Fairly poor</td>
<td>Small</td>
<td>Severe</td>
<td>No other Tx</td>
<td>1.17</td>
<td>0.56</td>
<td>96.34</td>
</tr>
<tr>
<td>55</td>
<td>Fairly poor</td>
<td>Very small</td>
<td>Severe</td>
<td>Other Tx avail.</td>
<td>1.08</td>
<td>0.51</td>
<td>96.84</td>
</tr>
<tr>
<td>56</td>
<td>Fairly poor</td>
<td>Small</td>
<td>Not severe</td>
<td>Other Tx avail.</td>
<td>1.04</td>
<td>0.49</td>
<td>97.34</td>
</tr>
<tr>
<td>57</td>
<td>Very poor</td>
<td>Very small</td>
<td>Severe</td>
<td>No other Tx</td>
<td>0.89</td>
<td>0.42</td>
<td>97.76</td>
</tr>
<tr>
<td>58</td>
<td>Very poor</td>
<td>Small</td>
<td>Not severe</td>
<td>No other Tx</td>
<td>0.85</td>
<td>0.41</td>
<td>98.16</td>
</tr>
<tr>
<td>59</td>
<td>Very poor</td>
<td>Very small</td>
<td>Severe</td>
<td>Other Tx avail.</td>
<td>0.76</td>
<td>0.37</td>
<td>98.54</td>
</tr>
<tr>
<td>60</td>
<td>Very poor</td>
<td>Small</td>
<td>Not severe</td>
<td>Other Tx avail.</td>
<td>0.73</td>
<td>0.36</td>
<td>98.90</td>
</tr>
<tr>
<td>61</td>
<td>Fairly poor</td>
<td>Very small</td>
<td>Not severe</td>
<td>No other Tx</td>
<td>0.67</td>
<td>0.34</td>
<td>99.23</td>
</tr>
<tr>
<td>62</td>
<td>Fairly poor</td>
<td>Very small</td>
<td>Not severe</td>
<td>Other Tx avail.</td>
<td>0.55</td>
<td>0.30</td>
<td>99.53</td>
</tr>
<tr>
<td>63</td>
<td>Very poor</td>
<td>Very small</td>
<td>Not severe</td>
<td>No other Tx</td>
<td>0.36</td>
<td>0.25</td>
<td>99.78</td>
</tr>
<tr>
<td>64</td>
<td>Very poor</td>
<td>Very small</td>
<td>Not severe</td>
<td>Other Tx avail.</td>
<td>0.23</td>
<td>0.22</td>
<td>100.00</td>
</tr>
</tbody>
</table>
expected worst case (constant), are in the ‘level of health improvement’, followed by changes in attribute levels for ‘value for money’, with change in ‘severity of health’ the next important, and change in ‘other treatments’ being the least important of the attribute-level differences. The largest single increment in utility is that from (fairly) poor value for money to (fairly) good value for money, with an increase in utility of 0.716.

The interpretation of the importance of each attribute by level is important relative to the reference case (constant). In terms of the changes in utility (logit function) for the attribute levels on severity (treating a severely affected patient group), while not being as large as that seen in some of the attribute levels for health improvement or value for money, it is seen to have an impact that is greater than some of the incremental impacts across differing levels of these attributes. For example, the impact of treating a severely affected patient group is seen to be greater than a change (a) from a ‘moderate’ health improvement to a ‘large’ health improvement or (b) from ‘fairly good’ to ‘very good’ value for money. The impact in the choice model of the attribute covering the availability of other treatments would appear to be small, although statistically significant. It is smaller than the other incremental changes in the model, although similar to the impact of moving from ‘fairly good’ value for money to ‘very good’ value for money.

The findings (Tables III and IV) show that the general public hold ‘severity of health condition’ and ‘the availability of other treatments’ as relevant social values, not to be ignored. However, the findings also indicate that in many instances the level of health improvement and value for money arguments may provide a strong indication of the social value (preference) associated with a health-care intervention.

At the level of the health technology scenarios, perhaps a more policy-relevant perspective, it is important to consider how the data presented can be interpreted in terms of the relative desirability (attractiveness) of the alternative scenarios. The logit scale, used in the utility functions, is linear in parameters (an additive model) and it provides values on an interval scale (interval scale properties) i.e. it is the difference between alternatives that is important. The data provide a relative measure of the desirability of the differences between alternative scenarios. It is not theoretically possible to interpret the utility values on a ratio scale, and it is therefore not possible to state how much one alternative may be regarded as better or worse than another. For example, the data show that the scenario ranked 10th is preferred to the scenario ranked 30th, and that the difference between the two scenarios in utility terms is 0.84, but it is not possible to quantify how much better (‘betterness’) the 10th-ranked scenario is compared with the 30th-ranked scenario.

The probability scale allows interpretation with ratio-level properties, and allows a judgment against a cardinal scale, indicating how much better or worse one option is compared with another, from the full factorial (64 scenarios). The probability values are arguably more suitable for use as an absolute measure of desirability across the alternatives (in the experimental design used). That is, when comparing estimates of the probabilities that each option will be chosen, it is possible to state how much worse (or better) one probability is compared with another. For example, using the probability scale it is possible to state that the scenario ranked 10th in Table IV is 2.33 times more likely to be chosen than the scenario ranked 30th in the table. Therefore, it is suggested here that the probability estimate can be
used as a measure of the ‘strength of preference’ across scenarios for the purpose of health-care decision making. For example, using the DCE probability data (Table IV) results suggests that 1 from any of the top 10 ranking health technology scenarios would be selected from the full factorial in around 40% of the cases (i.e. these technologies may be associated with a very strong public preference), and that in over 70% of the cases 1 scenario from the top 25 scenarios would be chosen. A number of scenarios are shown to be very unlikely candidates (i.e. a very weak public preference) for funding from the scenarios available (i.e. a less than 5% probability that 1 from bottom 10 scenarios would be chosen).

This notion of strength of preference is explored further elsewhere (Green, 2007, Chapter 8), where a range of health technology appraisal decisions documented by NICE have been used to demonstrate the potential value of the DCE data on probability, to inform on the social value of health-care interventions. The current paper is limited to the presentation of the methods used and the results from the discrete choice study. However, the findings presented are able to provide an important insight into the relationship between the social values used (as attributes) in the survey. The results support the existence of an equity versus efficiency trade-off, and allow some consideration of what might be meant by equity (in health technology appraisal) through the identification of specific trade-offs between the attributes, and combinations of attribute levels, used. The use of the DCE approach to estimate relative probabilities, across a set of plausible policy alternatives, is presented here as an extension of the more typical DCE analyses, as an application of DCE data as a means of exploring the general social value of health technologies.

The findings are presented as an aid to decision making, where specific information is already available, or judgments made, against the social value judgments (attributes) used in the survey. Assuming that attribute levels can be assigned to particular health technologies, in the context of a specific appraisal setting, findings from the DCE analysis can inform decision makers of the combinations of attributes and levels that may be more or less preferred at a societal level.

The study does have a number of limitations, reflecting the explorative nature of the research. The experimental design used is not complex, with a small factorial design used, and the results presented here are based on a simple analytical framework, using the conditional logistic model. Both of these factors are deliberate, to make the findings policy-relevant and the presentation of findings policy friendly, but represent potential limitations in the methodology. As an exploratory study we sought to keep the experimental design straightforward, but with feasibility established these preliminary findings can be used to inform more detailed future study designs. The current research indicates that the attributes used in the design were important, but it does not mean that other attributes could not be included in future studies of a similar nature. Pilot research did carefully consider a broad range of candidates for attributes in the survey design used, some of which may be worthy of further consideration in future studies, especially if using a broader priority-setting context. Such candidates may be the overall budget impact of health technologies or the number of people being treated. Both of these attributes were not directly relevant for the health technology appraisal setting used here, when applying the guidance for the NICE process (National Institute for Health and Clinical Excellence, 2005). Furthermore, including such attributes in a study presented to the general public would require research into the contextual nature of presenting such data to respondents. The attributes used covered issues that were expected to be very common in priority-setting dilemmas, such as technology appraisal. Other attributes could be introduced to make scenarios more context specific.

A further limitation is the absence of objective data on the response rate for the sample, due to the use of a random-location quota sampling approach. The sample used provides a good representation of the general population (Southampton), and while relatively small allows some suggestion of the generalisability of the findings. However, as this was an exploratory study, response rate would have been a good indication of the acceptability of the survey to respondents. On this issue, the feedback
from interviewers in the pilot study and the main survey was that respondents were keen to participate, seemed engaged in the survey, and generally had few problems completing the survey (accepting the fact that choices were difficult ones). That respondents reported some difficulty in answering the DCE choices suggests that respondents did engage with the choice context and suggests that they did weigh up the difficult choices presented, offering some confidence in the face validity of the experiment.

As with most empirical studies of this type, the study is also open to some level of criticism over the presentation, framing, and contextual approach adopted. However, it is suggested here that the results are useful and indicative of what may be possible in future, more comprehensive, research initiatives of this type. In its current form the preference data may be used, in a number of ways (together with other relevant information), to inform the difficult priority-setting dilemmas faced by health policy decision makers. Future research could address many of the limitations highlighted, for example, using qualitative methods to investigate the interpretation of attributes, terminology used, and the considerations when respondents make their choices.

7. CONCLUSIONS

This study adds to the sparse literature informing on the use of DCE methods to explore social preferences. The findings presented provide persuasive evidence on the feasibility and acceptability of the DCE approach for the elicitation of public preferences over priority-setting scenarios for health-care provision. It is the first such study to use the approach in a sample of the general public with the cost-effectiveness of alternatives reflected in the attributes of the DCE scenarios.

The current evidence base on what might be meant by equity is diverse and complex to interpret, and a number of commentators have highlighted the need for studies to consider the relationship between key social values. The study presented here has explored such relationships, and is able to provide useful insights into the relative importance of competing generic social value judgments, relevant for priority-setting decisions over health-care interventions.
### APPENDIX A

Characteristics of ‘social’ DCE studies (from Green, 2007)

<table>
<thead>
<tr>
<th>Study/context</th>
<th>Design (type/source)</th>
<th>Administration method and sample</th>
<th>Attributes (levels)</th>
<th>No. of choices per respondent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Farrar et al. (2000): Investigates preferences for clinical service developments</td>
<td>Fractional factorial used. 16 scenarios from possible 216 scenarios ((3^3 \times 2^1 \times 4^1 = 216)). Source of fractional factorial was software package SPEED</td>
<td>Self-complete questionnaire. Sample: UK Hospital Trust consultants ((n = 216); 130 respondents (60%))</td>
<td>Five attributes: level of clinical evidence ((A/B/C)), size of health gain ((\text{large/medium/small})), contribution to professional development ((\text{improve/no change})), contribution to education, training, and research contributes to ((0/1/2/all 3)), strategy area ((\text{none/local or national/local and national}))</td>
<td>Eight pairwise choices</td>
</tr>
<tr>
<td>Ratcliffe (2000): Social preferences over distribution of transplant organs</td>
<td>Fractional factorial used. Fractional design not stated. Full factorial of 108 scenarios ((3^3 \times 2^2)). Source of fractional factorial was software package SPEED</td>
<td>Postal, self-complete questionnaire. Sample: 800 randomly chosen university employees (UK) ((303 \text{ completed responses, } 38%))</td>
<td>Five attributes: age ((40/56/60)), alcoholic liver disease ((\text{yes/no})), expected length of survival ((5/10/15 \text{ years})), time spent on waiting list ((3/6/12 \text{ months})), re-transplanted ((\text{yes/no}))</td>
<td>Eight pairwise choices</td>
</tr>
<tr>
<td>Study/context</td>
<td>Design (type/source)</td>
<td>Administration method and sample</td>
<td>Attributes (levels)</td>
<td>No. of choices per respondent</td>
</tr>
<tr>
<td>---------------</td>
<td>---------------------</td>
<td>-----------------------------------</td>
<td>-------------------</td>
<td>-------------------------------</td>
</tr>
<tr>
<td>Bryan et al. (2002): Social preferences for different health-care programmes</td>
<td>Fractional factorial used, 16 scenarios: from a full factorial of 96 ((4^2 \times 3^1 \times 2^1)). Sixteen scenarios used in 24 choices. Source of fractional factorial was software package SPEED</td>
<td>Structured face-to-face interviews. Sample: General public (UK). Random sample of adults in health authority region; (n = 1762) invited, (n = 909) (51.6%) completed</td>
<td>Four attributes: number of people (1/10/100), chance of success (0.1/1/10/50%), survival (1/5 years), quality of life (for survey: EQ-5D descriptions, for analysis: 1.00/0.893/0.566/0.401)</td>
<td>Eight pairwise choices per respondent [three questionnaire formats used (randomly assigned)]</td>
</tr>
<tr>
<td>Gyrd-Hansen and Slothuus (2002): Preferences over willingness to forgo private consumption in order to obtain improved health-care services. Alternative health systems presented</td>
<td>Fractional factorial of 26 scenarios from full factorial of 800 ((2^5 \times 5^2)). Source of fractional factorial was software package SPEED</td>
<td>Interviews. Random sample, Danish population aged 20–74 years. (n = 1895), response rate of 69%</td>
<td>In summary, seven attributes [five with two levels (yes/no), two (money attributes) with five levels]: health system tries to offer all possible treatments irrespective of cost. More screening programmes introduced. Free choice of public hospital. Treatment in private hospitals is subsidised. Focus on preventative measures to reduce lifestyle-related diseases. Extra tax payment per year. Max out of pocket payment for health services</td>
<td>Three pairwise choices</td>
</tr>
</tbody>
</table>
Gyrd-Hansen (2004): Social preferences over health states (EQ-5D states)

Fractional factorial, using 42 scenarios (for 23 choices), from full factorial of 243 \((3^3 \times 2^3)\). Purposive selection of scenarios, using previous sample from a research study (MVH Study, York)

Interviews sample: random sample of the Danish population, \(n = 3,201\) (response rate of 49%)

Five attributes: used EQ-5D health descriptive system, five health dimensions with three levels each: mobility, self-care, usual activities, pain/discomfort, anxiety/depression

One pairwise choice.

Gerard (2005) (PhD thesis): Preferences on the principles that should be used to underpin service delivery, for out-of-hours (OOH) services

Fractional factorial, 18 scenarios. Full factorial of 81 scenarios \((3^3)\). Source: Statistical design catalogue. Used three different versions to explore impact of perspectives and compared results (personal versus social versus personal/social)

Self-complete postal survey. Sample: General public (UK), \(n = 798\) (response rate 36%) (non-random)

Four attributes: waiting time \((15/60/300\text{ min})\), fairness towards lower social class \((\text{none/limited priority/top priority})\), choice of doctor \((\text{none/two doctors/three doctors})\), extra taxation for OOH consultation \(\text{£0/£15/£75}\)

Nine pairwise choices

Baltussen et al. (2006): Social preferences of interventions, in priority-setting context

Fractional factorial, 24 scenarios. Full factorial of 64 scenarios \((6^3)\). Source: Statistical design catalogue

Not clear: self-complete (at workshop) or possible interview administration. Sample: policymakers (Ghana), convenience sample, \(n = 30\)

Six attributes: cost-effectiveness \((\text{yes/no})\), poverty reduction \((\text{neutral/positive})\), age of target group \((\text{young/adults})\), severity of disease \((\text{severe/not severe})\), health effects \((\text{high gain for few/low gain for many})\), total budget impact \((\text{high/low})\)

Twelve pairwise choices

Graded choice used (\% of budget), rather than discrete choices
<table>
<thead>
<tr>
<th>Study/context</th>
<th>Design (type/source)</th>
<th>Administration method and sample</th>
<th>Attributes (levels)</th>
<th>No. of choices per respondent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tappenden et al. (2007): Explores the social values used by the NICE technology appraisal process in the UK</td>
<td>Fractional factorial, 18 scenarios. Full factorial of 108 scenarios ( (3^3 \times 2^2) ). <em>Source:</em> SPSS ORTHOPLAN</td>
<td>Self-complete questionnaire. Sample: members of a health policy/decision-making body (UK), ( n = 37 ) (response rate 45%)</td>
<td>Five attributes: cost per QALY (( £15k/£25k/£35k )), uncertainty (low/high), age under (18/18–64/over 64), pre-treatment health status (0.25/0.50/0.75), availability of other therapies (yes/no)</td>
<td>Eighteen binary choices (yes/no)</td>
</tr>
<tr>
<td>Schwappach and Straßmann (2006): Social preferences in allocating budgets, primarily a ‘methods’ study</td>
<td>Fractional factorial used – 16 scenarios. Full factorial of 256 scenarios ( (4^3 \times 2^2) ). Source of fractional factorial was software package SPEED</td>
<td>Computer survey – web-based survey (interactive options) using a panel of respondents (Germany). Sample: convenience sample, public, ( n = 716 ) (response rate 72%)</td>
<td>Five attributes: age (child, teen, employable age, senior), QOL (low to low, low to high, mod to high, high to high), effect on life expectancy (loss of 5 years, no effect, gain of 5 years, gain of 10 years), frequency of disease (rare, common), cost of treatment (above ave., under ave.)</td>
<td>Four pairwise choices (plus three additional methods choices)</td>
</tr>
</tbody>
</table>
APPENDIX B

Showcard used to summarise the attributes used in the DCE survey

SEVERITY OF THE HEALTH CONDITION

Whether patients are severely affected by their condition

When considering severity we have judged that on at least one of the following areas patients have severe problems: (i) self-care (e.g. unable to wash or dress themselves), (ii) unable to perform usual activities (e.g. work, study, housework, family or leisure activities), (iii) extreme pain or discomfort, (iv) extreme anxiety or depression.

IMPROVEMENT IN HEALTH

Considering the average health improvement from treatment

Improvement in health refers to the benefits that the patient feels following treatment e.g. improvements in their mobility, improvements in their ability to perform usual activities, reduced pain, reduced anxiety.

In this questionnaire treatments offer one of the following:

- large improvement in health
- moderate improvement in health
- small improvement in health
- very small improvement in health

VALUE FOR MONEY

Whether or not the treatment offers ‘value for money’

‘Value for money’ is a common consideration within the NHS. Value for money is about the efficient use of resources (e.g. doctor’s time, hospital beds, healthcare funds).

In this questionnaire we use the following categories:

- very good value for money
- fairly good value for money
- fairly poor value for money
- very poor value for money

WHETHER OTHER EFFECTIVE TREATMENTS ARE AVAILABLE

Yes available or No not available

We assume that all patients will have usual and best supportive care available within the NHS (e.g. GP services, specialist outpatient appointments, and best supportive nursing care). Where referring to treatment options we are thinking about drugs, surgery, or specially developed services (such as physiotherapy, support services, occupational therapy, specialist education, preventative therapies, etc).
ACKNOWLEDGEMENTS

We would like to thank the Ipsos MORI Social Research Institute for undertaking data collection, especially Gregor Jackson. We would also like to thank Professor James Raftery, University of Southampton, for helpful comments. We would also like to acknowledge financial support from the Wessex Institute for Health Research and Development, University of Southampton, which funded data collection costs. Authors have no conflicts of interest to declare. At the time that the study was undertaken no ethics approval was required.

REFERENCES


