

Work in Progress: not to be quoted without author's permission

Paper for discussion at Health Economic Study Group Meeting, University of East Anglia, January 2002.

Listening to respondents: a think aloud study of Discrete Choice Experiment responses

Mandy Ryan¹, Caroline Reeves¹ and Vikki Entwistle²

**¹Health Economics Research Unit and ²Health Services Research Unit, University of
Aberdeen**

Abstract

The underlying assumption of all economic experimental (stated preferences) valuation techniques is that subjects have 'complete', 'stable', and 'continuous' preferences which can be elicited in a 'rational' manner. Following this a well-defined utility function can be estimated to calculate changes in welfare resulting from alternative health care policies. Utility theory is defined from a normative perspective. That is, it is not based on observations of how individuals behave, but rather on how they should behave. This paper presents results from a 'think aloud' study conducted to investigate the nature of people's thinking as they completed a discrete choice experiment. The aim was to investigate the economic axioms of completeness, continuity and rationality. Consideration was also given to understanding of cost and risk attributes within such an experiment. Preliminary results indicate that individuals are capable of forming preferences for unfamiliar commodities, apparently 'irrational' responses can be explained rationally and individuals often adopt non-compensatory decision-making strategies. There was evidence of cost-based responses and individuals had difficulty understanding risk.

Acknowledgements

We are grateful to Fernando San Miguel for allowing us to develop his questionnaire for this survey, to all respondents who completed the 'think-aloud' study, and to comments provided by Stirling Bryan and others at the recent MRC funded Workshop 'Qualitative analysis from the health economics perspective: meaning, best practice and future research', University of Bristol. Financial support from the Medical Research Council and the Chief Scientist Office of the Scottish Executive Health Department are acknowledged. The views expressed in the paper are those of the authors.

Background

Stated preference techniques are widely used in economics to elicit preferences for health care interventions. The underlying assumption of such techniques is that subjects have ‘complete’ and ‘stable’ preferences that can be elicited in a ‘rational’ manner. A further assumption is that individuals adopt compensatory decision making strategies (axiom of continuity). This implies that that when presented with valuation tasks respondents trade between the various dimensions of the scenarios presented (i.e. money, risk, time). Following on from these axioms, a well-defined utility function can be estimated for all individuals. From this function it is possible to estimate the change in welfare resulting from changes in, in the case of health economics, alternative health care policies.

Utility theory, summarised above, is defined from a normative perspective. That is, it is not based on observations of how individuals behave, but rather on assumptions about how they should behave. Whilst research has addressed issues such as the reliability, validity, and internal consistency of economic valuation techniques (Ryan et al, 2001), there is a notable lack of research investigating whether individuals satisfy the axioms of utility theory or understand the scenarios presented in valuation tasks, and thereafter, what the numbers generated from such studies actually mean. Tanur (1992) found that people often interpret survey questions in unexpected ways. Given the importance of developing valid methods of measuring benefits in health care, the meaning of the numbers generated from economic valuation tasks is a crucial area of future research. The importance of such research was also recently highlighted within the area of health status measurement (Mallinson, 2002).

Numerous qualitative research techniques have been used within areas such as psychology and sociology to investigate preferences and decision-making (Ryan et al, 2001). This paper applies one of those methods, the think-aloud technique, to investigate the nature of decision making within the context of a discrete choice experiment (DCE). The main aim was to investigate the whether individuals satisfy the economic axioms of completeness, rationality and continuity when they complete a DCE. Consideration was also given to understanding of the ‘cost’ and ‘risk’ attributes within the valuation task. The paper begins by providing background information on the three economic axioms of completeness; rationality and continuity have been tested to date. Consideration is also given to issues raised in the literature concerning the elicitation of willingness to pay and the value of risk reductions. Information on the think aloud method adopted in the study to explore these axioms is then provided. Preliminary results from the think aloud study are discussed. The paper concludes with some general points for discussion. This paper is very much work in progress, and all comments will be gratefully received.

Issues under consideration

Completeness of preferences

Experiments concerned with modelling individual preferences are based on the assumption of completeness of preferences. That is, when presented with information on a commodity, individuals can state their preferences for that commodity. If the commodity is new, or unfamiliar, then sufficient information must be given to respondents such that they can formulate their preferences before completing the valuation task (just as consumers formulate their values for newly marketed commodities).

A limited amount of work in the stated preference valuation literature in economics has attempted to test directly the completeness axiom. This has been mainly quantitative. Within the experimental WTP literature, environmental economists have developed models of ambivalence (Opaluch and Segerson, 1989; Ready et al, 1995), included ‘don’t know’ response options (Arrow et al, 1993; Wang, 1997; Lockwood, 1998), post-decision confidence measures (Li and Mattson, 1995; Hanemann and Kristrom, 1995) and ‘cannot answer’ responses (Common et al, 1997). There is some evidence that when presenting individuals such options as ‘don’t know’ or ‘neither’, subjects may choose such a response as an ‘easy’ option or for reasons other than their lack of preferences (Schuman and Presser, 1981; Oliver, 2000). Within the health economics an attempt has been made to overcome this by asking respondents to complete the same valuation task both within the same questionnaire as well as over time (San Miguel, 2000; Shiell et al, 1997; Shiell et al, 2000).¹ Such tests are limited in that they collect data on a subset of goods from the commodity space, and not the entire space. The use of qualitative work would improve on such tests

Within the psychology literature a distinction has been made between the *philosophies of articulated and basic value* (Fischhoff, 1991). The articulated values philosophy, which is consistent with economic theory, assumes individuals are able to articulate and express values on the most diverse topics. However, the basic values paradigm argues that people cannot be expected to have articulated opinions on more than a small set of issues of immediate concern. The only way people have well-formed preferences on a specific issue is by investing time and experience on that particular issue. This occurs only for a subset of goods. Thus, when preferences are valued for unfamiliar goods, people may “construct” their preferences as they complete the valuation task.

¹ Shiell et al (2000) also included additional qualitative questions which suggested that one third of respondents felt that the first interview had prompted them to think about their responses, and to change their responses in the second questionnaire. This figure fell to 24% when looking at preferences between the second and third questionnaire.

Continuity of Preferences

The continuity axiom assumes that individuals are willing to trade between the attributes presented in the valuation task– that is, they adopt compensatory decision-making. That is, for any pair of goods, people are willing to sacrifice units of one good for increases in the amount of other goods (unlimited substitutability). Preferences are represented in continuous indifference maps. However, non-compensatory theories of individual decision making have been proposed (Einhorn, 1970; Earl, 1983). Here the assumption of unlimited substitutability does not hold. Three types of non-compensatory (or discontinuous preference) models have been defined in the economics literature: lexicographic preferences; dominant attribute preferences and target setting (Scott, 2001). With lexicographic preferences the choice process assumes a complete hierarchy of preferences, whereby attribute 1, x_1 , is considered the most important attribute. That bundle which has the preferred level of x_1 will be preferred no matter what the level of other attributes. If x_1 is the same for two bundles then a second attribute, x_2 , is considered. This process continues for all attributes. When dominant attribute preferences exist, a ‘dominant’ attribute determines which alternative is chosen but there are no restrictions on the ordering or extent of trading for other attributes. In the case of target setting a certain level of a given attribute must be achieved before other attributes of the commodity are traded.

The dominance decision-making strategy has been tested for in the health economics DCE literature (Ryan et al, 2001; Scott, 2001). Such results are inconclusive since such choice behaviour may also be consistent with a twice differentiable utility function (Ryan and Gerard, 2001). This can be shown with the following simple example. Consider the responses to a DCE looking at two attributes: chance of successful outcome (0.3, 0.2, 0.1) and waiting time to operation (1 month, 2 month and 3 months). Suppose the following preference structure emerges:

$$(0.3,1)>(0.3,2)>(0.3,3)>(0.2,1)>(0.2,2)>(0.2,3)>(0.1,1)>(0.1,2)>(0.1,3)$$

Here the respondent always chooses the option with the highest chance of success. When this is the same, they choose the option with the lowest waiting time. Such a preference structure can also be represented by the following continuous utility function:

$$U(\text{success, waiting}) = 100\text{success} - \text{wait}$$

Whilst the respondent with the above preference structure has a strong preference for a successful outcome (as indicated by the parameter on the ‘success’ attribute), such a preference structure is also consistent with a continuous utility function. This does not suggest

that all preference structures are continuous, but rather that current continuity tests within DCEs are incapable of identifying violations of the continuity axiom. Given the possibility of identifying some continuous utility function for all possible responses, alternative continuity tests are required. Indeed, it is unlikely that quantitative analysis of responses to experiments can ever fully inform the debate on whether individuals conform to the axiom of *continuity* since all observed responses will be consistent with some utility function.

The assumption behind the economic models of decision-making is that individuals are utility maximisers. However, psychologists have argued that, given time, knowledge and computational limitations, respondents may employ simplifying heuristics (decision-making rules) (Kahneman et al, 1982), often employing ‘fast and frugal heuristics’ (Gigerenzer et al, 1999). Here the axiom of continuity often breaks down as individuals use minimum information strategies to make decisions. Examples of such strategies include the *recognition* heuristic where individuals adopt the simple heuristic of choosing the familiar option and *Take the Best* heuristic where individuals choose according to the level of one attribute, and if this is equal across choices they will move to a second attribute. This procedure continues across all attributes. This heuristic is similar to the lexicographic preference structure identified in economics. However, whilst economists claim such behaviour is utility maximising, psychologists claim it is a simplifying heuristic which may or may not be smart!

Rationality of Preferences

Rationality tests are included in stated preference exercises to ensure that the respondents are answering the choice tasks in a meaningful way. Rationality of responses have mainly been studied by including tests of dominance (or non-satiation) (Ryan et al, 2001), though transitivity and Sen’s alpha and beta properties have also been investigated more recently (McIntosh and Ryan, 2002; San Miguel, 2000).

McFadden (1999) argued that to define responses as truly ‘irrational’ it is necessary to collect additional information about respondents’ perceptions and beliefs, (which inform the decision process) as well as attitudes, motives and preferences. It is argued that when these elements are considered, apparent ‘irrational’ responses may be explained in a “rational” way. An example of this can be found in the WTP literature. Based on the definition of rationality used in WTP experiments (if commodity A is preferred to B, then individuals should be willing to pay more for A than B) evidence has been found of a ‘irrational’ responses i.e. respondents state a preference for one option and a greater WTP for their less preferred option. (Donaldson et al, 1997; Ryan and San Miguel, 2000). Based on respondent’s qualitative explanations of such responses, Ryan and San Miguel (2000) found that respondents providing “cost-based responses” could explain over 30% of apparently ‘irrational’ responses. Whilst such responses contradict the rationality criteria tested, they are

consistent with the 'fair price' decision making strategy defined in psychology i.e. people like to pay what they believe the cost is. They don't want to pay more and be exploited or pay less and be exploiting (Kahneman et al, 1986). There is evidence of this 'fair-price' behaviour in the provision of both private and public goods (Kahneman et al, 1986; Winer, 1986; Green et al, 1994). Similarly, when testing for transitivity, it has been shown in the literature that there are legitimate reasons for responses that fail to meet this axiom, including the tendency for respondents to focus upon particular features of the choice set (Fischer and Hawkins, 1993) and the attempt to reduce the psychological feeling of regret (Loomes et al, 1991; Loomes and Taylor, 1992). Indeed, better understanding of responses using qualitative methods has led to developments in Expected Utility Theory (EUT).

Willingness to pay

A number of concerns have been raised when eliciting willingness to pay for health care, including whether it is ethical to ask WTP within the context of health care and whether respondents are willing and able to understand the concept of maximum willingness to pay. One concern here is the evidence from the direct willingness to pay literature of cost based responses (Donaldson et al, 1997; Ryan and San Miguel, 2000). The study by Ryan and San Miguel (2000) attempted to limit cost-based responses in the experimental design by informing respondents in detail of the distinction between cost and value. Despite this, over 30% of respondents provided cost based responses. Schkade and Payne (1994), in using think aloud techniques to investigate how people respond to WTP questions, found that individuals justify their WTP responses by referring to the cost of the commodity being valued. The presence of cost-based responses raises concern for the use of willingness to pay within an welfare economics framework.

Understanding of risk

One area economists have long been concerned with is valuing reductions in risk. This is an important area in health economics, where many decisions are often made under uncertainty. There is an implicit assumption here that subjects understand the concept of risk, and respond to choice sets including risk in a rational manner. Important findings from the psychological literature challenge this assumption. It has been shown here that individuals have difficulty in understanding risk, and that they adopt simple decision-making heuristics. For example, individuals: view events as more likely if they are familiar; view hazards as more risky for others than themselves; respond to risk information differently if presented in terms of either gains or losses or as a relative risk compared to an absolute risk; and code risk data in a categorical manner i.e. 'low' or 'high' (Lloyd, 2001). Further, experimental studies have

shown that information about uncertainty and risks is better understood if presented in natural frequencies rather than in terms of probabilities and percentages (Gigerenzer et al, 1998).

Study Design

This study made use an existing *discrete choice experiment* (DCE) questionnaire and applied the *think aloud technique* to investigate adherence to the economic axioms of *completeness*, *rationality* and *continuity* as well as understanding of the *cost* and *risk* attributes.

Discrete choice experiment

DCEs are based on the premises that, first, any good or service can be described by its attributes and, second, the extent to which an individual values a good or service depends upon the nature and levels of these attributes. DCEs involve presenting individuals with choices involving scenarios described in terms of attributes and associated levels. For each choice they are asked to choose their preferred scenario. Response data are modeled within a utility function which provides information on: whether or not given attributes are important; the relative importance of these attributes; the rate at which individuals are willing to trade between attributes (marginal rate of substitution); and overall utility scores for alternative scenarios. If a price proxy is included as an attribute WTP can be indirectly estimated (Louviere et al, 2000).

The DCE approach was used for a number of reasons. Firstly, it may be argued that the choices posed in such an experiment reflect the type of decisions individuals make on a daily basis. Individuals very rarely think about their probability level of indifference between a gamble and a certain outcome (as required in a standard gamble experiment), the number of years they would be willing to give up at the end of their life for a better health state (as required in a time trade off experiment), nor the maximum amount of money they would be willing to give up for a health care intervention in a publicly provided health care system (as required in a WTP experiment). Adopting this format therefore reduced violation of the tests carried out because of a lack of understanding of the presentational format. Secondly, given the increased used of DCEs in health economics, it is important to have an understanding of how individuals interpret and respond to the choices within a DCE. Thirdly, this methodology requires subjects to give a direct preference for alternative bundles of goods. It therefore allows carrying out a direct test of the completeness axiom as defined in economic theory (see below).

A pre-existing DCE questionnaire concerned with eliciting preferences for screening for colorectal cancer. (San Miguel, 2000) was used. This questionnaire was chosen for a

number of reasons. Firstly the questionnaire included 4 attributes. If the continuity axiom is to hold within the framework of a DCE then it is more likely to do so when attributes are kept to a minimum level. If individuals do not trade between 4 attributes then they are unlikely to trade between a higher number. Secondly, the original DCE questionnaire was designed to test for completeness of preferences by repeating choices within the same questionnaire (see below). Whilst repeating choices within a questionnaire is potentially a useful method to test for completeness of preferences, potential limitations include respondents may simply recall their responses and the credibility of the experiment may be questioned if respondents feel they are being tricked. This study looked for evidence of this. Thirdly, the questionnaire included a cost attribute. Baron and Maxwell (1996) have argued that one way to attempt to overcome cost-based responses is to focus on the benefits from the commodities being valued (Baron and Maxwell, 1996). Given the experimental design of a choice experiment, where respondents look at the actual attributes of the commodity being valued, it has been hypothesised that DCEs may reduce the potential for cost based responses (Ryan and San Miguel, 2000). This study could test this hypothesis. Finally, risk was included as an attribute in this existing questionnaire, thus allowing consideration of the understanding of this attribute.

Following piloting of this questionnaire using personal interviews on 4 individuals a number of changes were made to the initial questionnaire in an attempt to clarify certain issues. This demonstrated the importance of conducting qualitative work in advance to ensure that respondents understand the questionnaire and choice tasks. However, to ensure comparability with the DCEs that have taken place to date no major changes were made to the questionnaire.

All participants were first given standard information about bowel cancer and screening tests. Following piloting of the questionnaire, respondents were told that: *'Bowel cancer (sometimes called colorectal cancer) affects women and men. It is the third most common type of cancer in the UK. It is more common in people over 40 than in younger people. The chance of a person getting bowel cancer at some point in their life is about 1 in 20. Bowel cancer can be cured if it is found early and treated properly. Screening tests aim to find bowel cancer in its early stages, before any symptoms appear. Most screening tests for bowel cancer are designed to test for invisible blood in the stool. The tests involve collecting samples of bowel movements with an applicator stick and smearing them onto special cardboard strips. The samples are sent to a pathology laboratory for investigation. Individuals are usually invited to have a test once every two years.'* Following the background information respondents were informed that programmes for testing for bowel cancer vary in lots of ways. They were then informed that in each of the 10 questions that follow (the DCE part of the questionnaire) they would be presented with 2 screening

programs for bowel cancer and asked which they would choose. Four attributes were systematically varied. The information provided on each of these attributes is shown in Table 1.

Table 1: Definitions of attributes included in the DCE

• Attribute in choice question	Definition of attributes provided in questionnaire
<ul style="list-style-type: none"> • How long people must restrict their diet for before having the screening test 	<p>Before taking the test, people should not eat or take:</p> <ul style="list-style-type: none"> - red meat (beef, lamb and processed meats) - raw fruits and vegetables (especially melons, radishes, turnips and horseradish) - more than 250mg of vitamin C (about 3 oranges) per day - some medicines such as aspirin or non-steroidal anti-inflammatory drugs (such as voltarin). <p>In our questions, the number of days for which people should not eat these foods or take these medicines before having the screening test varies between 1 and 3.</p>
<ul style="list-style-type: none"> • How likely it is that the tests will suggest that someone has cancer when in fact the person does not. 	<p>With any screening test, there is a chance that the test will suggest that someone has a disease when in fact they do not. When a screening test mistakenly suggests that someone has cancer, the result is said to be a 'false positive' result. 'False positive' results are a problem because they cause people to worry and possibly to be treated unnecessarily.</p> <p>In our questions, the screening tests that we ask you to think about give false positive results to either 8, 15 or 20 out of every 1000 people who are screened (that is 0.8%, 1.5% and 2% of people who are screened).</p>
<ul style="list-style-type: none"> • The cost of the test 	<p>Although in reality you would NOT be asked to pay for such a test, for the purposes of the questionnaire you are asked to imagine that you would have to pay. This can help us to find out how much you would value the screening tests.</p> <p>In our questions the costs of the screening tests are £16, £32 or £48.</p>
<ul style="list-style-type: none"> • Whether people are told when their test results are normal 	<p>Sometimes, health services only tell people the results of their screening tests if the screening tests find a problem.</p> <p>In our questions, the screening programmes we ask you to consider will either write to everyone within three weeks whether the result is positive or negative, OR will only write if the test result is positive and will tell them to assume that the test is negative if they don't hear.</p>

Experimental design software was used to create the choice set (San Miguel, 2000). With the inclusion of quantitative completeness, dominance and transitivity tests (see below) each

respondent was presented with 10 choices. For each choice they were asked to choose screening programme A, B or neither. They were informed that the neither option implied that they would prefer not to be screened (following the pilot where it was found that neither was interpreted as preferring an alternative option that had been presented in the questionnaire. Figure 1 shows an example of the choices presented to respondents, and Table 2 shows the entire choice set.

Figure 1: Example of choice question from DCE

Q2 Which screening programme would you choose for yourself?
Please tick one of the boxes below

PROGRAMME:	A	B	
Number of days of dietary and medication restrictions	1	2	
Number of people who will get a false positive result	15/1000	15/1000	
Cost of the test	£16	£48	
Will I be told if I have a normal test result?	No	Yes	
I would choose	A <input type="checkbox"/>	B <input type="checkbox"/>	Neither <input type="checkbox"/>

Table 2 Difference between choices in DCE

Choice	Number of days of dietary and medication restrictions	Number of people who will get a false positive result	Cost of the test	Will I be told if I have a normal test result? (0=no, 1=yes)
1	2	12/1000	£32	-1
2*	1	0	£32	1
3	-1	-12/1000	£0	1
4	0	5/1000	-£16	1
5**	-2	0	£32	-1
6*	1	0	£32	1
7***	1	7/1000	£16	0
8***	1	0	£32	0
9**	-2	0	£32	-1
10	0	-5/1000	£16	1

* Completeness *Test1*, ** Completeness *Test2*, *** Dominance tests

Think aloud technique

To explore the economic axioms and understanding of the cost and risk attributes the think aloud technique was used (Ericsson and Simon, 1984; Carroll and Johnson, 1990). There are

two types of think aloud: concurrent and retrospective. It is recommended that a combination of the two be used wherever possible (Ericsson and Simon 1993) Concurrent think aloud is where people are asked to verbalise what they are thinking as they complete a certain task. Retrospective think aloud asks people to describe what they were thinking after the task has been completed. During the pilot work three types of the think-aloud technique were tested. Initially respondents were asked to think-aloud but to expect interruptions from the interviewer. The interviews were carried out with a lot of probing and examination during the think-aloud, as this seemed, to the interviewer, to be the most natural and convenient way to conduct the sessions. However, it was felt that this methodology may be interfering too much with the thought processes of the respondent and could possibly be leading, and therefore introducing bias. A second approach was therefore investigated: respondents were requested to think-aloud for choices, with no interruption from the interviewer (except for reminders to keep talking out loud). This proved to be a difficult task, with respondents finding it difficult to keep talking and expressing concern about repetition. Long periods of silence also suggested that not all thoughts were being conveyed. Following this a third method was tested. Here respondents were asked to think-aloud for two or three choices, and then to reflect back after this. It was felt that this method would not interfere too much with thought processes, but would allow the interviewer to explore fully how people went about making their choices. Based on the recommendations of the interviewer following the pilot survey, the third approach was adopted. Thus, an attempt was made to employ both concurrent and retrospective methods. Instructions were clearly provided by the researcher. Participants were instructed to state out loud everything that went through their minds as they were completing the choices. They were asked not to explain or plan out what they were thinking, but to act as if they were speaking to themselves. If the respondent was silent for a period of time s/he was reminded to keep thinking aloud. Before performing the think aloud task a practice session was conducted. This was a game of noughts and crosses.

There was no coaching on how to answer the choices initially. However, the research team agreed that if respondents asked for clarification then further explanation would be provided. As they were completing the questionnaire the interviewer asked additional questions and prompted for clarification if necessary.

Sample and setting

A purposeful sample of nineteen men and women (ten men and nine women) aged between 49-69 years were recruited from two bowling clubs in Aberdeen, Scotland (given that if screening were to be offered, this is the likely age group). The bowls clubs received £20 for every individual who took part. All interviews were conducted by one of the authors (CR) at the University of Aberdeen.

Method of analysis

Face-to-face interviews were tape-recorded and later transcribed. The charting approach was used to analyse the transcripts. Here a chart was produced to record references to the three economic axioms of interest as well as to the cost and risk attributes. The transcripts were reviewed and comments related to these themes recorded on the chart.

In addition to the qualitative data, the original questionnaire included quantitative tests of completeness and rationality (dominance) of preferences. These were maintained in the study. The completeness tests employed compared the responses provided in repeated choices *within* the same questionnaire. Two tests (*Test1* (choices 2 and 6 in Table 2) and *Test2* (choices 5 and 9 in Table 2)) were included, thus allowing the ‘construction of preferences’ hypothesis to be tested. Four response patterns were possible from the two repeated choices: satisfaction of *Test1* and *Test2* indicates complete preferences; failure of both tests incomplete preferences; failure of *Test1* and passing *Test2* indicates construction of preferences or learning and satisfaction of *Test1* and failure of *Test2* random error. The power of these completeness tests depend on the nature of the repeated choices i.e. it had to be ensured that they were not ‘too easy’ to pass. Both tests involved trading for at least three of the four attributes. Where responses differed for the repeated choice the interviewer asked the respondent to elaborate on the reasons for their repeated response. Consideration was also given to whether respondents commented on the fact that the questionnaire included repeated choices. A further direct was included: at the end of the choices with subjects being presented with the following statement: ‘If I was to ask you to start again now (I won’t!) would you go about it in the same way, or might you try to think about them differently?’

The original *rationality* tests were 2 dominant choices (choices 7 and 8 in Table 2). Failures of these tests were identified and reasoning for choices made investigated. It is recognised that such tests may challenge the credibility of the valuation task (since they are often obvious!). This was explored in the transcripts.

Results

General issues regarding the use of the think aloud technique are being presented in a separate paper which is currently being prepared. This paper focuses on results concerning the three axioms and understanding of the cost and risk attributes. However, it is worth noting that respondents had difficulty in concurrently thinking aloud.

“I find it a bit harder to tell you as I’m reading”

Thinking aloud retrospectively appeared to be easier and the interviewer, on numerous occasions, had to remind the interviewee to think aloud.

To date 7 transcripts have been provisionally analysed so results reported are first thoughts. All 7 transcripts need to be re-read (limited time has meant that they have not been read in sufficient detail) as well as a further 12 analysed. Despite this, a number of points have emerged that will be further explored in further analysis. The results from the Charting exercise are shown in Appendix 1. What follows are some initial interpretations of these results.

Completeness of preferences

Respondent 1 showed some signs of preference formation. For example, in the early choices he states that he would not consider paying £38, but by the end of the ten choices is willing to pay this amount. He also recognises that if he was given the questionnaire to do a second time he may change his answers:

‘Possibly, possibly, if I were going to take this away with me and read what I’ve just wrote on there, read through it again right? There is a possibility that I might change something’

However, a number of other respondents indicated, from the direct question, that they would not want to change their responses if they completed the questionnaire a second time. Whilst this provided initial support to the axiom of completeness, it may also be the case that respondents felt some pressure to say this. This raises the general question of the influence of interviewer presence when conducting think aloud studies. However, conclusions regarding completeness of preferences were supported by satisfaction of *Test1* and *Test2*.

Continuity of Preferences

There is strong evidence of non-compensatory decision making in responses, with individuals often using the information on one or two key attributes to make their decisions. Some evidence is provided here of individuals adopting ‘the take the best’ heuristic.

‘always uppermost in my mind was the failure ratio so I always went for the failure ratio irrespective of error cost. But when the failure was given let’s say it was 20:20 the dietary didn’t mean anything more to me. I mean I would have gone in any sort of diet if told so but ah when the failure sort of rate error was the same error I then pointed towards going for the most expensive test because I felt it would probably well it must be a better test otherwise it would cost the same’.

However, there was also evidence of individuals trying to be ‘consistent’ and ‘correct’ in the way that they responded. The easiest way to be consistent would be to choose according to the level of one attribute, and, only if this is equal across choices, move on to the second attribute. Thus, respondents may have been more concerned with answering consistently for the interviewer than with maximising utility. For example, respondent 1 said:

‘I’ll tick that. I hope I’m doing it correct really’

This would also explain high rates of satisfaction of the completeness tests.

Rationality of Preferences

No respondent commented on the credibility of including dominant choices, though one respondent did comment on the ease of answering such a choice. Whilst a number of respondents ‘failed’ the dominance tests, analysis of responses indicate that failures may be a result not of misunderstanding the choice task, but rather of making assumptions about the levels of attributes which the researcher may not have taken on board. For example, respondent 2 failed the dominance test in Choice 8 because he assumed that a higher cost was to be preferred (see cost based responses below).

‘I would lean towards the cost of a test at £48 being you’d think you’d get a more thorough test.’

Similarly, respondent number 4 ‘failed’ Choice 7 because he sees more days of dietary restrictions as to be preferred, arguing that this will result in a better chance of getting a result.

‘No, I’ll go for option B again on that one 2 days. I think it’s again it would be giving you a bigger chance of getting err perhaps getting the result, a better result.’

This also happened for Choice 8. These initial results suggest that more consideration should be given to respondents who are defined as ‘failures’ when using the simple dominance test of rationality.

Willingness to pay

Whilst one individual clearly understand the concept of cost representing value, another did not consider cost on the basis that this should not have to be considered in the NHS. There

was evidence of cost-based responses. For example, when answering Choice 6, respondent 4 stated:

'And £48 per one test as opposed to £16 err seems to me as though you must be getting a more thorough test other wise there wouldn't be such a difference In cost so in 6 I would be going for B.'

And, respondent 7 states more generally:

'So the reason I'm not considering cost is because it's not possible for me to judge it in the total balance of the spends of the National Health Service.... If bowel cancer costs £2 million for every person then spending £100 per test might be worth spending'

Understanding of risk

Within this study the risk variable (number of people who will get a false positive result) was presented in natural frequencies (though these were also converted to percentages for respondents). Despite this, respondents had difficulty in understanding this attribute. For example, respondents 4 and 6 both perceive 20/1000 false positives as better than 8/1000.

Discussion

Stated preference techniques are widely used in economics to elicit preferences for health care interventions. Considerable resources have been devoted to exploring methodological issues in the application of these techniques (Ryan et al, 2001). Such research has been mainly quantitative in nature. Whilst this quantitative research has been crucial in the development of experimental valuation techniques, consideration must also be given to whether individuals satisfy the axioms on which these techniques are based and whether they understand the attribute definitions. Only if they do can the numbers produced be interpreted as meaningful. Quantitative analysis can go some way to addressing these issues. However, the various qualitative techniques that are used within disciplines such as sociology and psychology are likely to be very useful in this research agenda. This paper presented one such method, the think aloud technique, and reported preliminary findings from a study which applied this technique to investigate responses to an existing DCE questionnaire.

To date only 7 transcripts have been analysed, and this analysis has been preliminary (unfortunately we ran out of time!). It is clear that respondents found it difficult to think aloud. Given the potential wealth of information that can be generated from qualitative think

aloud studies, future work should explore methods of improving respondent's ability to think aloud.

Preliminary results from the charting exercise suggest that individuals are capable of forming preferences for unfamiliar commodities, apparently 'irrational' responses can be explained rationally and individuals often adopt non-compensatory decision-making strategies. There was evidence of cost-based responses and individuals had difficulty understanding risk. It is recognised that, given the study group were elderly people, future research should investigate whether there are differences between different socio-economic groups in terms of decision-strategies, rationality, completeness of preference, understanding of attributes etc. Future work should also investigate the extent to which the results apply to other commodities (e.g. more familiar versus less familiar commodities) as well as other valuation techniques (standard gamble, time trade off and willingness to pay).

Whilst data analysis is at a very early stage, it is clear that useful information can be derived from listening to respondents for the conduct and design of future DCEs, and indeed valuation tasks more generally. Individuals, when thinking about the price proxy within a DCE, tend to think about cost rather than value. One way to overcome this problem may be to move away from defining the price proxy in terms of cost. Future work should explore alternative definitions of the price proxy. Whilst the psychological literature has indicated that individuals find proportions easier to handle than probabilities when considering risk, this study found individuals had difficulties understanding this attribute. Future work must address the issue of how best to convey risk to respondents in valuation tasks.

Whilst non-compensatory decision-making does not appear to be the approach taken by the majority in this study, questions are raised concerning whether respondents were maximising their chances of being consistent or utility? It is possible that individuals adopted simple decision making heuristics to help them appear consistent to the interviewer. This would also explain the relatively good results for the completeness and rationality tests, as well as the direct question on changing responses. There was some evidence of this. For example, respondents 4 said:

'I'll tick that. I hope I'm doing it correct really'.

This raises general questions concerning how to conduct think-aloud studies, with the potential problems of interviewer bias (especially when adopting an approach where the interviewer intervenes regularly, as in this study).

Related to compensatory decision making, there was some evidence of 'Take the Best' decision heuristic being adopted. If individuals do adopt this simple heuristic, is it a smart heuristic? Does it maximise utility (and therefore is it simply a lexicographic preference

structure) or does it give a second best solution? The work by Gigerenzer et al (1999) has argued that there exists simple decision making heuristics that 'make us smart'. The implication here is that using simple decision making strategies individuals can proxy optimal outcomes. However, these heuristics have been mainly tested within the context of inferences rather than preferences. For example, when asked which is the largest city, respondents use the recognition heuristics i.e. they choose the one that they have heard of. It is a simple matter to show that this is generally a smart heuristic. When looking at inferences there is usually some gold standard against which the decision heuristic can be compared to assess whether it is smart. Within the area of preference elicitation it may not be so easy to observe the gold standard i.e. the utility maximising outcome. This is a crucial area for future research and suggestions for experiments, potentially laboratory based (?), would be welcome.

References

- Arrow, K., Solow, R., Portney, P., Leamer, E., Radner, R. and Schuman, H. Report of the NOAA panel of contingent valuation. *Federal Register*, 10, 4601-14, Washington DC, 1993.
- Baron, J. and Maxwell, N. Cost of public goods affects willingness to pay for them. *Journal of Behavioural Decision Making* 9, 173-183, 1996.
- Carroll, J.S. and Johnson, E.J. *Decision Research: A Field Guide*. Sage, Newbury Park, 1990.
- Donaldson, C., Shackley P. and Abdalla, M., Using willingness to pay to value close substitutes: carrier screening for cystic fibrosis revisited, *Health Economics* 6, 145-159, 1997.
- Earl P. *The Economic Imagination*. *Wheatsheaf books*, Brighton, 1983.
- Einhorn H.J. The use of nonlinear, non-compensatory models in decision-making. *Psychological Bulletin*, 221-230, 1970.
- Ericsson, K.A. and Simon, H.A. *Protocol Analysis: Verbal reports as data*. MIT Press, Cambridge, 1984.
- Fischer, G.W. and Hawkins, S.A., Strategy compatibility, scale compatibility, and the prominence effect, *Journal of Experimental Psychology-Human Perception and Performance* 19, 580-597, 1993.
- Fischhoff, B. Value Elicitation: Is there anything in there? *American Psychologist* 46, 835-847, 1991.
- Gigerenzer G, Todd P and the ABC Research Group. *Simple Heuristics that Make us Smart*. New York: Oxford University Press, 1999.
- Gigerenzer G. Hoffrage U. Ebert A. AIDS counselling for low-risk clients. *AIDS Care*, 10, 197-211, 1998.
- Green, D., Kahneman, D., and Kunreuther, H. How the scope and method of public funding affects willingness to pay for public goods. *Public Opinion Quarterly* 58, 49-67, 1994.
- Hanemann, W. M. and Kristrom, B. Preference uncertainty, optimal designs and spikes, in *Current Issues in Environmental Economics*, Johansson, P-O. Kristrom, B. and Maler, K-G (eds), New York, Manchester University Press, 1995.
- Kahneman D, Slovic P, Tversky A. *Judgement Under Uncertainty: Heuristics and Biases*. Cambridge, United Kingdom: Cambridge University Press, 1982.
- Kahneman, D., Knetsch, J.L. and Thaler, R., Fairness as a constraint on profit seeking. Entitlements in the market, *American Economic Review* 76, 728-741, 1986.

- Lloyd, A. 2001. The extent of patients' understanding of the risk of treatments. *Quality in Health Care* (forthcoming).
- Li, C-Z. and Mattsson, L. Discrete choice under preference uncertainty: An improved structural model for contingent valuation. *Journal of Environmental Economics and Management*, 28, 256-269, 1995.
- Lockwood, M. Integrated value assessment using paired comparisons. *Ecological Economics*, 25, 73-87, 1998.
- Loomes, G., Starmer, C. and Sugden, R., Observing violations of transitivity by experimental methods, *Econometrica* 2, 425-439, 1991.
- Loomes, G. Taylor, C., Non-transitive preferences over gains and losses. *Economic Journal* 102, 357-365, 1992.
- Louviere, J. J., Hensher, D. A. and Swait, J. D. Stated choice methods. Analysis and Application. Cambridge, *Cambridge University Press*, 2000.
- Mallinson S. Listening to respondents: a qualitative assessment of Short-Form 36 Health Status Questionnaire. *Social Science and Medicine*, 54, 11-21, 2002.
- McFadden, D. Rationality for economists? *Journal of Risk and Uncertainty* 19, 73-105, 1999.
- McIntosh E. Ryan M. Using discrete choice experiments to derive welfare estimates for the provision of elective surgery: implications of discontinuous preferences *Journal of Economic Psychology* (forthcoming)
- Oliver A. Complete preferences over health states: a reply to the paper by Shiell et al. *Health Economics*, 9, 47-55, 2000.
- Opaluch, J.J Segerson, K. Rational roots of 'irrational' behaviour. New theories of economic decision making, *Northeastern Journal of Agricultural Resource Economics*, 18, 81-95, 1989.
- Ready, R.C, Whitehead, J.C. and Blomquist G.C. Contingent valuation when respondents are ambivalent. *Journal of Environmental Economics and Management*, 29, 181-196, 1995.
- Ryan M. Gerard K. Using choice experiments to value health care programmes: current practice and future challenges. Paper presented at International Health Economic Association Conference, University of York, July 2001.
- Ryan M. San Miguel F. Testing for consistency in willingness to pay experiments. *Journal of Economic Psychology* 21, 305-317, 2000.
- Ryan M. Scott D. Reeves C. Bate A. van Teijlingen E. Russell E. Napper M. and Robb C. Eliciting public preferences for health care: a systematic review of techniques. *Health Technology Assessment Report*, 5(5), 2001.
- San Miguel F. 2000. Testing the assumptions of completeness, stability and rationality of preferences using discrete choice experiments. PhD Thesis, University of Aberdeen.
- Schuman H. and Presser S. Questions and answers in attitude surveys: Experiments on Question Form, Wording and Context, Academic Press, New York, 1981.
- Shiell, A., Hawe, P. and Seymour, J. Values and preferences are not necessarily the same. *Health Economics Letters*, 6, 515-518, 1997.
- Shiell, A., Seymour, J., Hawe, P. and Cameron, S. Are preferences over health states complete? *Health Economics*, 9, 47-55, 2000.
- Schkade, D., and Payne, J., How people respond to contingent valuation questions: A verbal protocol analysis of willingness to pay for an environmental regulation. *Journal of Environmental Economics and Management* 26, 88-109, 1994.
- Scott A. Identifying and analysing dominant preferences in discrete choice experiments: an application in health care. *Journal of Economic Psychology* (forthcoming)
- Tanur, J.M. *Questions and answers*. New York: Russell Sage Foundation, 1992.
- Wang, H. Treatment of "Don't Know" responses in contingent valuation surveys: a random valuation model. *Journal of Environmental Economics and Management*, 32, 219-232, 1997.
- Winer, R. A reference price model of brand choice for frequently purchased products. *Journal of Consumer Research*, 13, 250-256, 1986.

Appendix 1 Preliminary Results from Charting Exercise

	Completeness	'Rationality'	Continuity/Decision-Making Heuristics	Cost attribute	Risk attribute
R1	<p>Q2 =A, Q6=Neither; Q5=A, Q9=A</p> <p>Initially would not consider paying £38, but by choice 10 willing to pay this amount.</p> <p><i>Change answers if go back?</i> Possibly, possibly, if I were going to take this away wi' me and read what I've just wrote on there, read through it again right? There is a possibility that I might change something'.</p> <p>No mention of repeated tests</p>	<p>Pass both</p> <p>No mention of credibility</p>	<p>Appeared to be considering levels of all attributes when making decision i.e. adopting compensatory decision-making. 'So basically I've said that, that, that and that, so I'd go there., A. I would choose A.</p>	<p><i>Payment vehicle</i> – understood in terms of 'imagine going private.' – suggested this himself'.</p> <p>Understands distinction between cost and value – '...and then when it comes to this one, cost of the test, now you've changed it again from 32 to 48 now it all boils down to this fact that the cost, as I say, if you've got to pay for something, ken it's how much of a value do you put on this screening..'</p>	<p>Understands that 15/1000 is better than 20/1000</p>
R2	<p>Q2 =B, Q6=B; Q5=A,Q9=B</p> <p>Justifies Q5 on the basis that cheaper and Q9 on the basis of higher cost therefore better quality.</p> <p><i>Change answers if go back?</i> 'I think I'd stick with the answers I'd put down because I did look at them and err make up my mind for one reason or the other and it was quite definite what I felt ...'</p> <p>No mention of repeated tests</p>	<p>Pass Q7</p> <p>Fails Q8 - going for most expensive test 'I would lean towards the cost of a test at £48 being you'd think you'd get a more thorough test.'</p> <p>No mention of credibility</p>	<p>Based responses on level of 'false positives' and 'cost' (preferring the most expensive test, assuming it is a better quality).</p> <p>Evidence of 'Take The Best' or lexicographic decision making heuristic 'always uppermost in my mind was the failure ratio so I always went for the failure ratio irrespective of err cost. But when the failure was given let's say it was 20:20 the dietary didn't mean anything more to me. I mean I would have gone in any sort of diet if told so but ah when the failure sort of rate err was the same err I then pointed towards going for the most expensive test because I felt it would probably well it must be a better test otherwise it would cost the same'.</p>	<p><i>Evidence of cost based responses consistently:</i> for example, in choice 2 'It would be a more positive sort of test if its, it's only a third of the, if it's 3 times more than £16 it's going to be more positive test I would imagine so I'd go for it'; Choice 4: Well in number 4 I would then, I would there again go for A. er because, because there is less chance of a wrong result than you have in B and the cost of £32 opposed to £16 so there must be a reason for that difference in cost means that er you must be getting a better examination. Choice 6: 'And £48 per one test as opposed to £16 err seems to me as though you must be getting a more thorough test other wise there wouldn't be such a difference in cost so in 6 I would be going for B.'</p> <p>Tries to rationalize why cost varying across choices</p>	
R3	<p>Q2 =B, Q6=B; Q5=A,Q9=A</p> <p><i>Change answers if go back?</i> I would think so, I mean there's only 1 way is there?</p> <p>No mention of repeated tests</p>	<p>Both correct</p> <p>No mention of credibility</p>	<p>Individual is choosing according to one attribute – whether told if normal test. Seems to be heavily influenced by personal experience of daughter in law from whom information was withheld on her cancer. Does not adopt compensatory decision-making.</p>		

	Completeness	'Rationality'	Continuity/Decision-Making Heuristics	Cost attribute	Risk attribute
R4	<p>Q2 =B, Q6=B; Q5=A,Q9=A</p> <p><i>Change answers if go back?</i> 'After changes in initial choices because did not understand risk states that, in response to question would you answer them in the same way 'More or less. I think I would after the changes in the first one or two.'</p> <p>No mention of repeated tests</p>	<p>Fails Q7 - sees 2 days dietary restrictions as better than 1. 'No I'll go for B again on that one 2 days. I think it's again it would be giving you a bigger chance of getting err perhaps getting the result a better result.'</p> <p>Fails Q8 - 'Well again I'd go for B in that one the 3 days again clearing the system and giving you a better chance to get a result. I'll go for B there.'</p> <p>Generally, assumes 3 days dietary requirements better than 1</p> <p>No mention of credibility.</p>	<p>Individual choosing according to 'false positives' – dominant attribute (but does not understand it).</p>	<p><i>Cost based responses:</i> Choice 2 – 'I think that if you're spending more on the test it would be more thorough perhaps.No I'll stick with actually B in the question 2. The cost I think you'd be getting a better test'</p>	<p>'<i>Understanding of risk</i>' – individuals misunderstands the risk attribute, seeing 20/1000 as being preferred to 8/1000. At Choice 4 the interviewer explains this and goes back to Choice 1.</p>
R5	<p>Q2 =A, Q6=A; Q5=B,Q9=B</p> <p>No mention of repeated tests</p>	<p>Both correct</p> <p>Comments on Q8 – 'It's an easy answer again it's A. It's so easy for me as its got the best result although it's a small percentage again but 8 out of a 1000 and I'm getting a result as I say I'm not worrying about cost but the costs less so it turns into an easy answer for me.</p>	<p>Evidence of non-compensatory decision making. – 'whether told' was a key attribute, then seems to consider other attributes. Explaining Choice 8: 'Well A because although I said earlier on I wouldn't worry about the cost but the cost is less and the number of days and the restriction in your medication and diet is less and the other 2 answers are similar so it would have to be B.'</p>		
R6	<p>Q2 =A, Q6=A; Q5=B, Q9=B</p> <p>'No I think I would try and answer them in the same way.'</p> <p>No mention of repeated tests</p>	<p>Fails Q7 – perceives $15/1000 > 8/1000$ 'Well it's the false positive is hard in question A compared to B and I think it's the more people that's told the better,... That's why I chose B.</p>	<p>Individual initially choosing according to number of days of dietary and medication restrictions. In response to being asked why choose A in Choice 1, said 'the fact of the diet'. And in explaining choosing A in choice 2: 'Well I think it because it's over the period of the 24 hours sort of thing. I'd be more inclined to stick to the diet for the one day rather than the two.</p> <p>By Choice 7 is showing some evidence of compensatory decision making – Explaining Choice 7 – 'I'd be inclined to go with two days for more people to be told than just one day and just half the people to be told.'</p>	<p>Does not consider cost at all: 'But as I say cost shouldn't come into it anyway in the National Health as it is'</p>	<p>Misunderstanding of risk: in Choice 4 the respondent is asked to think about risk (since dietary restrictions are the same and does not consider cost or whether told. Once interviewer explains, respondent chooses A in Choice 4. Yet by Choice 7 misunderstands again and fails dominance check because believes 15/1000 is to be preferred to 8/1000.</p> <p>In choice 10 justifies A in Choice 10 on basis that same number of dietary restrictions but higher number of false positives (since thinks this is a good thing).</p>
R7	<p>Q2 =B, Q6=B; Q5=Neither,Q9=Neither</p> <p>'I don't think so...'</p> <p>No mention of repeated tests</p>	<p>Q7 seems to be easy to pass ' So I think, following on A is obviously the one which I'd. Since I'm not really looking A is obviously a much better deal than B in question 7.</p> <p>Q8 – Neither</p>		<p>Cost based responses: 'So the reason I'm not looking at the cost is because it's not possible for me to judge it in the total balance of the spends of the National Health Service. ...if bowel cancer costs £2 pounds for every person then spending £100 per test might be worth spending...</p>	