ARTICLE

Using discrete choice experiments to value health care programmes: current practice and future research reflections

Mandy Ryan[1] and Karen Gerard[2]

[1]Health Economics Research Unit, University of Aberdeen, UK; [2]Health Care Research Unit, University of Southampton, UK

Abstract: There has been growing interest in discrete choice experiments (DCEs) in health economics over the last few years. This paper identifies and describes applications conducted during 1990–2000. From this basis some important areas for future research are reflected upon. These include: having a better understanding of how respondents interpret price, risk and time attributes; strengthening designs and analysis; investigating decision making heuristics employed when completing DCEs, and the extent these are related to the complexity of the task; external validity, reliability and generalisabilty. Collaborative work with statistical design experts, psychologists, sociologists and qualitative researchers will prove useful when investigating these issues. It is also critical to link, more generally, the research agenda to work being carried out in marketing, transport and environmental economics and health economics benefit assessment.

Keywords: discrete choice experiments, review, stated preference, methodology

Introduction

To fully assess the value of alternative health policies in public health care systems, researchers must estimate the value of non-marketed commodities. Several non-market stated preference valuation techniques have been used in the health economics literature, and over the last few years there has been a growing interest in discrete choice experiments (DCEs). The aim of this paper is to identify current practice in the application of DCEs in health care and, from this, reflect on important areas for future research. A brief introduction to DCEs including guidance on conducting and reporting DCEs is presented. It is, however, not the intention of this paper to describe how to conduct a DCE; for more on this, as well as detailed explanation of the theoretical basis of DCEs, see Louviere et al (2000). The health economics applications reviewed (see Appendix 1) and the methods for identifying and reviewing applications of DCEs in health economics are explained. Many of the issues raised in this paper cannot be discussed in detail due to limited space; references are provided for the interested reader.

Discrete choice experiments

Discrete choice experiments draw upon Lancaster's economic theory of value (Lancaster 1966) and random utility theory (McFadden 1973, p 105–42; Hanemann 1984). Attributes of the commodity being valued are first defined, and levels assigned to them. Statistical design theory is then commonly used to draw an independent sample of scenarios (ie combinations of attribute levels) from the full factorial set. These are placed into efficient choice sets and subjects are presented with these choices. Assuming that price is included as an attribute in these choice sets (or some proxy for price), willingness to pay (WTP) can be indirectly estimated for both changes in individual attributes, as well as changes in any combination of attributes.

The decision making process within a DCE can be seen as a comparison of indirect utility functions. The subject makes a series of choices. For each choice he/she chooses the alternative that leads to the higher level of utility. Thus, if:

$$U_{iq}(A) = vi_q(A) + \varepsilon_{iq} \qquad (1)$$

where $U_{iq}(A)$ represents the indirect utility function of individual q for good i with attributes A, $v_{iq}(A)$ represents the measurable component of utility estimated empirically, with i, q and A as defined above, and ε reflects the unobservable factors, the subject will choose i over j if:

$$(v_{iq} + \varepsilon_{iq}) > (v_{jq} + \varepsilon_{jq}) \qquad (2)$$

or

$$(v_{iq} - v_{jq}) > (\varepsilon_{jq} - \varepsilon_{iq}) \qquad (3)$$

Correspondence: Mandy Ryan, Health Economics Research Unit, University of Aberdeen, Foresterhill, Aberdeen AB25 2ZD, Scotland, UK; tel +44 1224 554 965; fax +44 1224 550 926; email m.ryan@abdn.ac.uk

Given that error terms are unknown, a probability model is estimated where:

$$P(i_q \mid A, C) = Pi_q = [(\varepsilon_{jq} - \varepsilon_{iq}) < (v_{iq} - v_{jq})] \qquad (4)$$

The probability of choosing alternative i (over j) by individual q, given the set of attributes A and the choice set C, is given by the probability that the error difference is smaller than the difference in the 'observable' utility component between i and j. For purposes of empirical measurement, a probability distribution is assumed for $(\varepsilon_{jq} - \varepsilon_{iq})$. Logit and probit techniques are commonly used to estimate the measurable component of the utility function. Given that multiple observations are obtained from individuals (since individuals are presented with a series of choices) panel data models are appropriate.

When analysing the data, assumptions must be made about the functional form of the observable indirect utility function. Assuming a linear utility function (as is most commonly done), then:

$$V = \chi_0 + \Sigma \chi_n X_n + \delta P + \theta \qquad (5)$$

where χ_0, χ_n and δ are the parameters of the model to be estimated; X_n's represent the levels of the n attributes of the commodity being valued ($n-1,2...,k$); $\Sigma \chi_n X_n$'s represent the summation of all the model effect coefficients; P the price level (or some proxy for price); and θ the unobservable error term for the model. χ_0 reflects the subject's preferences for one commodity over another when all attributes in the model are the same (referred to in the literature as the alternative specific constant (ASC)).

The χ parameters are equal to the marginal utilities of the given attributes (ie $\partial v/\partial X_k = \chi_k$), and the ratio of any two parameters show the marginal rates of substitution between attributes. Following on from this, the ratio of any given attribute to the absolute parameter on the price attribute shows how much money an individual is willing to pay for a unit change in that attribute (ie χ_n/δ) (McIntosh and Ryan 2002, p 376). From this it is possible to estimate overall WTP for a change in the provision of a service by summing all attributes coefficients (ie $\Sigma \chi_n X_n$) and dividing by the absolute parameter on the price – the measure commonly used to estimate the welfare impact of a change in policy is given by:

$$\text{WTP} = \{(\Sigma \chi_n X_n)/\delta\} \qquad (6)$$

Methods

A systematic literature search was conducted to identify published English language studies using DCEs within a health economics context.[1] Searches of the following databases between 1 January 1990 and 31 December 2000 were undertaken: MEDLINE; EMBASE; HEALTHSTAR; Social Science Citation Index; PsychLIT; EconLIT; and Health Management Information Consortium database. To reflect the different terms used to refer to DCEs, search strategies were formulated for each individual database using the following free text terms: 'conjoint', 'conjoint analysis', 'conjoint measurement', 'conjoint studies', 'conjoint choice experiments', 'part-worth utilities', 'functional measurement', 'paired comparisons', 'pairwise choices', 'discrete choice experiments', 'discrete choice modelling', 'discrete choice conjoint experiments' and 'stated preference'. Studies were also identified from the bibliographies of DCE applications identified in the literature search and websites of academic institutions (the latter to identify discussion papers and published conference papers). Only experimental (as opposed to actual) studies that used choice-based response data, were written in English and applied to health care (as opposed to methodology, theory or opinion papers) were of interest.

When conducting a methodological systematic review, issues are raised concerning the criteria that should be used to evaluate quality. This was done by considering current practice regarding the key stages of a DCE (selection of attributes and assignment of levels, choice of experimental design and construction of choice sets, measurement of preferences, estimation procedure and testing of validity of responses). In addition, the papers reviewed were read to identify criteria assessed, and current debates in the literature concerning appropriate criteria to assess instruments were considered (Ryan et al 2001, p 17–23). Based on this, studies were examined with respect to 26 'items' which were both extractable from secondary sources and presented a good breadth of coverage of issues in the conduct of DCEs (see Tables 1–5). Twenty-four items reported counts of occurrences or facts to classify current practice, retaining as much objectivity as possible. The authors used their own judgements' to classify the remaining item (Table 3, item 13, 'quality' of experimental design).

Results

The literature search generated 919 references which were checked for duplicates and ineligibility. Revealed preference studies, conjoint analysis studies not grounded in random utility theory (RUT) and studies not written in English or without health economic applications were excluded. This reduced the pool of potential studies to 328. A second, more careful reading of abstracts removed a further 199

references because they were not original empirical studies. One hundred and twenty-nine full references were obtained and read by one of the authors (KG) to verify eligibility. A further ninety-five studies were rejected as they did not use choice-based responses[2] eg, ranking exercises (Singh et al 1996), rating exercises (Ryan et al 1998) or a choice-based approach using an alternative modelling framework (Hadorn et al 1992)[3] The remaining thirty-four studies were included in the review. Appendix 1 provides key references relating to eligible studies[4]

Table 1 provides background details of the studies used in the review. The UK was the predominant producer of DCEs during the decade covered and the rate of production of studies increased over time.

Although the main focus was to value benefits, other applied areas included health insurance planning, labour supply characteristics, time preference values and valuation of generic health status domains.

Attribute selection and level setting

Table 2 considers current practice with respect to attribute selection and level setting. Whilst it is not possible to comment on the appropriateness of the sources used to select attributes and assign levels, it was encouraging that all studies reported sources and discussed the rationale for them.

The number of attributes included varied from 2 to 24, with a mode of 6. No studies assessed the effect of the number of attributes on respondent's ability to complete the choice task, though there appears to be an assumption that

Table 1 Background to DCEs

Item	Category	Nr studies
1. Country of origin	UK	20
	US	7
	Australia	6
	Canada	1
2. Publication year[a]	1990–1995	5
	1996–1997	3
	1998	6
	1999	7
	2000	13
3. Areas applied	Economic evaluation	17
	Insurance plans	5
	Other (eg labour supply)	12
4. Preference source	Patient	10
	Community	11
	Health insurance consumer	5
	Other (eg Consultants, GPs)	8

[a] If more than one publication was associated with a study the earliest one was counted.

Table 2 Attribute selection and level setting

Item	Category	Nr studies
5 and 6. Source of attributes and level assignment	Clearly described	34
7. Number of attributes	2–3	5
	4–5	10
	6	9
	7–9	4
	10	2
	> 10 (max 24)	4
8. Attributes covered[a]	Monetary measure	19
	Time	25
	Risk	12
	Health status domains	19
	Health care	28
	Other	3

[a] Totals more than 34 as each study may use many attributes.

4–6 attributes is acceptable (and beyond that the choice task will get too complex). Louviere (2001, p 33–4) argues that increasing the number of attributes will not significantly affect mean preference parameters. He comments that there is no empirical evidence to suggest this but that increasing numbers of attributes (and other aspects of complexity) would impact on the random component variability. Hensher et al (2001, p 374) note, whilst academics and practitioners agree that DCEs should not be too 'complex', to date there is no guidance on what constitutes 'complex'. This is clearly an area for future research.

The scope of attributes used in DCEs is clearly linked to the topic of application. Many studies used DCEs as a way of valuing non-health benefits and 19 studies (58%) incorporated a monetary attribute. Whilst including a price proxy allows indirect estimation of willingness-to-pay (WTP), several issues are raised (Ratcliffe 2000). One concerns the appropriate payment vehicle in a collectively funded health care system. Whilst most studies used payment at the point of consumption, willingness to accept compensation (Study 14, see Appendix 1), travel costs (Study 20) and Medicare Levy (Study 24) were also used as payment vehicles. Two studies (Studies 31 and 32), concerned with incentives for changing general practitioner (GP) behaviour, used changes in income levels. Future applications of DCEs to collectively funded health care interventions should systematically explore alternative 'incentive compatible' payment vehicles.

Another important area for future research is the extent to which DCEs overcome some of the known problems of the direct WTP approach eg, lack of scope sensitivity, part-whole biases, status quo bias, cost-based responses, strategic biases, 'yea-saying' and warm glow (Ryan et al 2001).

Whilst it has been argued DCEs can overcome some of these problems (Hanley et al 2001), as yet there has been no empirical work in this area.

Twenty-five (74%) studies included some type of time attribute. Sixteen of these incorporated waiting time, whilst other definitions of time included travel time, time to return to normal activities, duration of illness and time preferences. Study 20 estimated value in terms of time and from this a monetary value was estimated using the value of waiting time for public transport. Given that time is a commodity that individuals are used to trading in a publicly provided health care system, its potential use within an economic evaluation should be explored. However, a number of caveats should be noted, including the fact that the value of time depends on the opportunity cost of that time and waiting time (to treatment) may have an impact on the risk of worsening outcomes.

Twelve DCEs (35%) included risk as an attribute in a variety of formats; three presented risk levels in terms of gains; five as loses; and four were mixed. Information was presented as percentage chances (9 studies), natural frequencies of chances (2 studies) or both (1 study), and as absolute risk (5 studies) and relative risk (6 studies). It has been shown in the psychological literature that individuals have difficulty in understanding risk, viewing events as more likely if they are familiar; hazards as more risky for other people than themselves; responding to gains or losses and relative risk compared to absolute risk differently; and code risk data in a categorical manner ie, 'low' or 'high' (Lloyd 2001). There appeared to be little consideration given to whether these studies had appropriately presented risk as an attribute. Future work should address this.

One study was identified which had addressed the sensitivity of WTP estimates to the range of the attribute levels (Study 33). This study found that whilst estimated coefficients were not significantly different across five of the six attributes included in the experiment, mean WTP estimates were significantly different for four of the five welfare estimates. Within the marketing literature, Ohler et al (2000) found that whilst attribute range influences main effects to a small degree, substantial effects were found on attribute interactions and model goodness of fit. More empirical data would provide a stronger basis to investigate this issue in greater detail.

Experimental design and choice sets

Table 3 reports current practice regarding experimental designs employed and the construction of choice sets. Inclusion of interaction terms must be allowed for at the design stage of the study. The first aspect examined was whether fractional or full factorial designs were used. A full factorial design is more robust as it allows investigation of all interaction effects. But the price paid is potentially large numbers of scenarios to be examined by respondents. Thus, full factorials are usually only possible if there are a small number of attributes and levels. Four studies used small full factorial designs (Studies 3, 6, 15 and 22), 25 employed a fractional factorial design, often allowing only for main effects, and 5 did not specify. Despite evidence from outside health economics that main effects explain over 80% of the preference structure (Emery and Baron 1979; Louviere 1988, p 95; Pearmain et al 1991, p 37) (health) economic theory has little guidance for researchers on what interactions to include. Indeed, it is likely that the importance of interaction terms will be highly context specific. Whilst this review did not judge the relevance of the main effects designs in individual studies, the tendency to include only main effects mirrors practice in the environmental economics literature. However, here the view is changing and it would be apposite for health economists to keep apace with these developments (Adamowicz 2002).

Table 3 Experimental design and construction of choice sets

Item	Category	Nr studies
9. Design type	Full factorial (main, interactions, polynomials, higher order)	4
	Fractional factorial (limits on effects estimated)	25
	Not reported	5
10. Design source	Software packages	19
	Catalogue	2
	Expert	4
	Not reported, fractional factorial design used	4
	Not reported, no further details on design	5
11. Design plan	Main effects only (includes interactions with covariates)	25
	Main effects and selected 2-way interactions between attributes	2
	Full factorial (all effects, includes higher polynomial)	4
	Not reported	3
12. Allocation of scenarios to choice sets	Binary	2
	Pairwise (random)	15
	Pairwise (other)	9
	>2 choices	5
	Not reported	3
13. Design quality	'Strong'	0
	'Weak'	21
	Not reported	13

Of the 25 studies making use of fractional designs 19 used computer software packages to produce the design, 2 used design catalogues and 4 utilised design experts. The same 5 studies as above did not report on design source. Making no details available about design type and/or source is poor practice, preventing the reader or end user making judgement on whether an optimal design had been used. A word of caution is also needed regarding design software packages and design catalogues. It would be too simplistic to unthinkingly use these without examining the properties of the resultant designs. Either method may contain irregular fractions through error or design faults. Simple checks (correlation matrix of the design's attribute levels, frequency count of levels) should be conducted.

The majority of studies within health economics have used computer software to design DCEs. Such software, like catalogues, produces a number of scenarios for which preferences must be elicited. If using the single binary choice approach, ie would you do this, 'yes' or 'no' (eg Studies 19 and 30) then the scenarios derived from the experimental design can be presented to individuals. However, when multiple choices are used (as in at least 29 studies in the review) the resulting scenarios from the experimental design must be placed into choice sets. In doing this, it is important that the statistical design properties of orthogonality, balance, minimum overlap and balanced utilities are maintained (Huber and Zwerina 1996). Zwerina et al (1996), recognising that it is difficult to satisfy all four design principles, have developed a computerised approach which minimises the size of the covariance matrix.

To what extent have these properties been considered in the application of DCEs in health economics? A number of (earlier) applications paired scenarios randomly, making little reference to design properties. Fourteen studies used a constant comparator (from the experimental design or the current situation), aiding adherence to the statistical properties of computer generated and catalogue-based designs. Study 20 produced a design where the three attributes, 'cost', 'waiting time' and 'location' were entered into the computer software design as 'cost in choice A', 'location in choice A', 'waiting time in choice A', 'cost in choice B', 'location in choice B' and 'waiting time in choice B'. This resulted in a computer generated design that was automatically orthogonal in differences (since the level for say location in A was compared to location in level B, and the difference was therefore automatically given). Such an approach is only possible for experiments with a small number of attributes (since the number of attributes is

doubled in the design). One study used computer software to develop choices (Hanley et al 2002) (Study 25). Five studies used more than two scenarios in choice sets (Studies 2, 13, 18, 23, 25) and used appropriate allocation methods (ie L^{MN} designs or random designs) (Louviere et al 2000, p 111–37). Three studies failed to report information about the allocation of scenarios to choice sets.

This review graded designs 'strong' if they clearly reported on three or four of the principles of a good design, and discussed the suitability of the resultant design; and 'weak' if the study has not been explicit in this regard but some information can be extracted indirectly (eg a full factorial will be orthogonal and balanced). Based on this, there were no strong designs and 21 weak designs. Thirteen studies (38%) did not report sufficient detail to comment on design. Whilst there is no one correct way to design DCEs, greater attention should be given to reporting the properties of designs. Greater collaboration with design experts would help to improve designs and make use of cutting-edge techniques.

Measurement of preferences, response rate and comprehension

Table 4 summarises current practice in the measurement of preferences and respondent comprehension. Thirteen studies used 8 or less choices, eighteen studies between 9 and 16 choices and 2 studies used more than 16 choices. One study did not comment on this issue. Whilst no research has investigated what a reasonable number of choices is in health economics, this has been debated elsewhere. Hanley et al (2002), within the environmental economics literature, found that increasing the number of choice tasks influenced estimated model parameters. However, Hensher et al (2001), within the context of choices over airlines, concluded the opposite. This is clearly an important area for future research.

Most studies (27) used self-complete questionnaires (posted or handed out and returned in a 'clinic-like' setting), 3 administered surveys by interview, and 3 used computerised interviews (one did not report information). Whilst it is generally accepted that interviewer-based administration is the best approach, this is also a more costly approach and may not be feasible within the researcher's budget or timetable. Care must also be taken to understand the impact of the computerised interview approach on design properties; at present this does not appear to be well understood.

Most self-completion questions used text, with little consideration given to pictorial representations of the subject matter. Given current developments in high-powered graphic

Table 4 Measurement of preferences and comprehension

Item	Category	Nr studies
Measurement of preferences		
14. Number of choices per respondent	8 or less choices	13
	9–16 choices	18
	More than 16 choices	2
	Not clearly reported	1
15. Administration of survey	Self-complete questionnaire	27
	Interviewer administered	3
	Computerised interview	3
	Not reported	1
16. Postal reminders	None	6
	1 reminder	6
	2 reminder	6
	Not reported	9
Comprehension		
17. Response rate	90% – 100%	7
	70% – <90%	7
	50% – <70%	7
	30% – <50%	6
	<30%	2
	Not reported	5
18. Ease of task as perceived by subject	10% – <20% of subjects had difficulty	6
	20% – 35% of subjects had difficulty	4
	Not reported	24
19. Completion time	10 –15 minutes	3
	16 – 60 minutes	1
	1– 2 hours	1
	Not reported	29

packages and the ability to download images from the Web, consideration should be given to alternative styles of presenting health goods to the subject. More generally, consideration should be given to whether the type of representation (text, pictorial, interview (with or without pictures), computer-based (with or without pictures) influences the model parameters. In the marketing literature Vriens et al (1998) suggest that whilst pictorial representations improve respondents' understanding of the design attributes, verbal representations facilitate judgement.

Response rates varied across studies. The seven studies achieving 90%–100% response rates handed out questionnaires in 'clinic-like' settings or were interview-based. Lower responses were obtained using postal questionnaires. Many of the very low response rates were reported in pilot studies. It is unacceptable not to report response rates (of which 5 studies did not). Ten studies explicitly asked respondents to comment on how easy they found the questionnaire to complete. This provides some helpful, but limited, direct feedback. There did appear to be a significant minority who felt they had some difficulty,

but given the nature of the question it was not clear whether this arose from the choosing per se or because the instructions were unclear. Where it was measured, completion time to fill out questionnaires seemed mostly acceptable (three studies took on average 10–15 minutes to complete) but completion times of 1–2 hours seemed too long. Future research should address the issue of comprehension within the context of alternative survey designs (ie varying number of attributes, levels and presentation of scenarios (text versus pictures).

Estimation procedure

Table 5 reports the use of econometric models to analyse DCE data. The majority of studies (26) employed random effects probit or logit models, thus allowing for multiple observations from respondents. Where respondents are asked to evaluate more than two alternatives within a choice set multinomial logit had been most commonly employed. This approach assumes independence of irrelevant alternatives and does not allow for multiple observations. The random parameter logit, employed in one study (Study 25), overcomes these limitations, whilst also allowing for heterogeneity of responses (Revelt and Train 1998). Other methods being developed within environmental economics to model DCEs, allowing for heterogeneity of responses, include mixed logit models, latent class models and hierarchical Bayes models (Adamowicz 2002). These models will become increasingly important to health economics as research attempts to understand the complexity of the choice process within a DCE (see below).

Validity and reliability

Table 5 also reports the different aspects of validity considered. Given that many of the applications have taken place in countries with a publicly provided health care system, and therefore with a lack of secondary data sets to compare real and stated behaviour, much emphasis is placed on theoretical validity (22 studies or 65%). This involves checking that model coefficients have the signs expected given theory. A more indirect test of theoretical validity is to consider whether results are consistent with intuition ie, whilst economic theory has nothing to say about the relationship between experience and value, we may expect that patients who have experience of an illness value treatment differently from those with no experience. Such 'intuitive hypotheses' will be context specific. Two studies addressed face validity (presenting a pre-pilot version of the survey to a subset of respondents and revising it to

remove ambiguities contained in the interpretation of the tasks and information provided), and 4 studies assessed convergent validity (tested with respect to WTP, standard gamble and visual analogue scale).

Utility theory is defined from a normative perspective. That is, it is not based on observations of how individuals actually behave, but rather on how they should behave. Whilst some argue that economics should be concerned only with prediction, others argue that it is important to test validity by looking at the extent to which individual behaviour is consistent with the axioms of economic theory. This latter view has lead to the validity of DCEs being tested by investigating rationality of responses (often referred to as internal consistency) and the assumption that individuals trade across all attributes (equivalent to the Archimedian (and related continuity) axiom).

Fifteen studies tested for rationality by including dominance tests (choice sets where one alternative is clearly superior) and three studies tested transitivity. Noticeable here was a lack of any research investigating reasons for 'irrational responses'. There is a growing literature from both psychology and economics indicating that apparently 'irrational' responses can be rationally explained (Fischer and Hawkins 1993). Future work should explore 'irrational' responses in more detail and qualitative research techniques will prove useful here (Schkade and Payne 1994). Alternative more challenging tests of rationality may be explored, including transitivity and Sen's contraction and expansion properties (Sen 1993). Applications of transitivity tests would imply observing the following choice pattern: if Choice A is preferred to Choice B and Choice B is preferred to Choice C, then Choice A should be preferred to Choice C. See

McIntosh and Ryan (2002) for an application of this test to DCEs. Applications of Sen's contraction (expansion) property imply that if Choice set 1 is narrowed (expanded) to form Choice set 2, and the alternatives chosen from 1 are still in 2, then no unchosen alternative should be chosen and no chosen alternative should be unchosen. See San Miguel (2000) for an application of these tests to DCEs.

One of the underlying assumptions of DCEs is that individuals adopt compensatory decision making. That is they consider all the attributes included in the experiment and, based on the levels of all attributes, make a choice. Twelve studies investigated this assumption by testing whether individuals always chose according to the best level of a given attribute. However, such tests are limited since compensatory behaviour may result in such a choice pattern (Ryan and Gerard 2001). Indeed, quantitative analysis is limited in its ability to identify decision strategies employed by respondents and, again, qualitative work will prove useful in this area (Ryan et al 2002).

Interestingly, there has been little debate regarding the relationship between adherence to economic axioms and complexity of the DCE design. Whilst a limited number of economists have discussed this (De Palma et al 1994; Heiner 1983; Scott 2002), more extensive literature from psychology has argued that as complexity of tasks increase, so would reliance on simple decision making strategies (Payne 1976; Payne et al 1993; Gigerenzer and Todd 1999, p 141–67). This may lead to violations of the assumptions of rationality and compensatory decision making and is an essential area for future research. Further, if non-compensatory decision making is identified, questions are raised concerning the analysis of response data. Recently Swait (2001) has developed a formal model for analysing a wide range of decision strategies that relax the assumption of compensatory decision making. Future work should consider the application of this model in health care.

No studies were identified addressing external validity, ie comparisons of hypothetical and actual behaviour. Only one external validity study was identified in the environmental economics literature (Carlsson and Martinnsson 2002). Here DCE estimates of values for wildlife protection were insignificantly different from real payment obtained in an experimental setting. Health economists must attempt to apply tests of external validity (as they are doing in the direct WTP literature (Blumenschein et al 2001; Clark 2002). Consideration should be given here to setting up experiments to compare real and stated behaviour, or identifying secondary data sets where information on actual behaviour exists.

Table 5 Estimation procedure and validity

Item	Category	Nr studies
Estimation procedure		
20. Econometric models	Probit/ordered probit	3
	Random effects probit	17
	Logit/MNL (multinomial logit)	9
	Other – random parameter logit, OLS, weighted least squares	3
	Not reported	2
21–26. Validity and reliability[a]	Face	2
	Theoretical	22
	Convergent	4
	Rationality	18
	Compensatory decision making	12
	Reliability	1

[a] Totals more than 34 as each study may test for more than one aspect of validity.

Table 6 Reflections on future research questions

Design questions
- What is the manageable number of attributes and choices to include in a DCE?
- If indirect WTP measures are to be extracted, to what extent can DCEs overcome known problems with the direct WTP approach?
- Under what circumstances is time a useful valuation rubric?
- How can risk attributes be meaningfully incorporated into DCEs?
- To what extent do attribute range effects exist?
- Can more complex experimental designs improve the precision of DCE modelling?
- Can good practice criteria be developed to promote strong quality experimental designs?

Analysis questions
- How do alternative approaches to econometric modelling compare?
- Can analysts be more discerning in their treatment of apparently 'irrational' responses?
- To what extent should respondents conform to the axioms of economic theory?
- Are DCEs applied in health economics externally valid and reliable and under what circumstances are results generated generalisable?

One study (Study 11) investigated the test-retest reliability of DCEs. Whilst a high level of reliability was reported, the authors note that their results are likely to be context specific, and that further work is required in this area.

Generalisability

In addition to the criteria set for this review, generalisability ought to be considered. Preference elicitation can be time-consuming and expensive, resulting in an increasing interest in developing generalisable benefit measures (eg EQ-5D has been used to measure the value of alternative health outcomes across a wide range of health care interventions). Whilst information on this issue is currently lacking within the DCE literature (Dowie 1998), a useful starting point would be to build on the *benefit transfer* literature in environmental economics (Morrison et al 2002). Benefit transfer relates to the notion of assessing whether DCE models can be transferred or adapted to construct valuations for resources that are different in type, location or time from the one originally studied (Smith 1993).

Conclusions

Discrete choice experiments are being increasingly used in health economics in a number of different contexts. The growing enthusiasm by health economists to use DCEs must not be overlooked by the need to follow a paced

development that both draws from, and builds on, knowledge from other disciplines. This paper has reviewed current practice in the application of DCEs in health economics, and reflected on a number of important design and analysis issues for future research (see Table 6 for a summary). It is important to note that there are no cookbook answers to many of the research questions raised. Consideration must be given to the context of the design. It is clearly also critical to link this research agenda to valuation work being carried out in health economics more generally.

Acknowledgements

Notes

1 Published papers included published conference proceedings with ISBN numbers.
2 Although it is possible that ranking and rating data can be converted to choice-based responses, the studies cited here did not adopt that approach.
3 This framework was a modification of Thurstone's method of analysis of pairwise choices which allowed for a degree of preferences.
4 In instances where a number of papers were associated with a single study the study was deemed eligible if at least one reference was published between 1st January 1990 and 31st December 2000.

Appendix 1

Studies reviewed (for full references see Ryan and Gerard 2001)

Study 1   Propper C. 1990. Contingent valuation of time spent on NHS waiting list. *Econ J*, 100:193–9.
Propper C. 1994. The disutility of time spent on the United Kingdom's National Health Service waiting lists. *J Hum Res*, 30:677–700.

Study 2   Chakraborty G, Ettensen R, Gaeth G. 1994. How consumers choose health insurance. *J Health Care Market*, 14:21–33.

Study 3   Ryan M, Farrar S. 2000. Using conjoint analysis to elicit preferences for health care. *BMJ*, 320:1530–3.

Study 4   Ryan M. 1999. Using conjoint analysis to take account of patient preferences and go beyond health outcomes; an application to in vitro fertilisation. *Soc Sci Med*, 48:535–46.

Study 5   Ryan M, Hughes J. 1997. Using conjoint analysis to assess women's preferences for miscarriage management. *Health Econ*, 6:261–73.

Study 6   Lee A, Lum AE, Beehan SJ, Hillman K. 1996. Inter hospital transfers: decision-making in critical care areas. *Crit Care Med*, 24:618–22.

Study 7   Farrar S, Ryan M, Ross D, Ludbrook A. 2000. Using discrete choice modelling in priority setting: an application to clinical service departments. *Soc Sci Med*, 50:63–75.

Study 8   San Miguel F, Ryan M, McIntosh E. 2000. Demonstrating the use of conjoint analysis in economic evaluations: an application to menorrhagia. *Appl Econ*, 32:822–33.

Study 9   Stensrud J, Sylvestre E, Sivadas E. 1997. Targeting Medicare consumers. Managed care providers can make inroads by understanding preference and cost-sensitivity issues. *Market Health Serv*, 17:8–17.

Study 10 Ryan M, Bate A. 2001. Testing the assumptions of rationality, continuity and symmetry when applying discrete choice experiments in health care. *Appl Econ Lett*, 8:59–63.

Ryan M, Bate A, Eastmond C, Ludbrook A. 2001. Using discrete choice experiments to elicit preferences. *Q Health Care*, 10:155–60.

Study 11 Bryan S, Buxton M, Sheldon R, Grant A. 1998. Magnetic resonance imaging for the investigation of knee injuries: an investigation of preferences. *Health Econ*, 7:595–603.

Bryan S, Gold L, Sheldon R, Buxton M. 2000. Preference measurement using conjoint methods: an empirical investigation of reliability. *Health Econ*, 9:385–95.

Study 12 Freeman JK, Szeinbach SL, Barnes JH, Garner DD, Gilbert FW. 1998. Assessing the need for student health services using maximum difference conjoint analysis. *J Res Pharm Econ*, 9:35–49.

Study 13 Szeinbach SL. 1998. Using maximum difference conjoint and visual analog scaling to measure patients' utility for a particular health state. *J Res Pharm Econ*, 9:83–100.

Szeinbach SL, Barnes JH, McGhan WF, Murawaski MM, Corey R. 1999. Using conjoint analysis to evaluate health state preferences. *Drug Inf J*, 33:849–59.

Study 14 van der Pol M, Cairns J. 1998. Establishing patient preferences for blood transfusion support: an application of conjoint analysis. *J Health Serv Res Policy*, 3:70–6.

Study 15 Vick S, Scott A. 1998. Agency in health care: examining patients' preferences for attributes of the doctor–patient relationship. *J Health Econ*, 17:587–605.

Scott A, Vick S. 1999. Patients, doctors and contracts: an application of principal-agent theory to the doctor–patient relationship. *Scott J Polit Econ*, 46:111–34.

Study 16 Scott A, Watson MS, Ross S. 2003. Eliciting preferences of the community for out of hours care provided by general practitioners: a stated preference discrete choice experiment. *Soc Sci Med*, 50:804–14.

Scott A. 2002. Identifying and analysing dominant preferences in discrete choice experiments. *J Econ Psychol*, 23:383–98.

Study 17 Cunningham MA, Gaeth GJ, Juang C, Chakraborty G. 1999. Using choice-based conjoint to determine the relative importance of dental benefit plan attributes. *J Dent Educ*, 63:391–9.

Study 18 Hakim Z, Pathak D. 1999. Modeling the EuroQol data: a comparison of discrete choice conjoint and conditional preference modelling. *Health Econ*, 8:103–16.

Study 19 Lowin A, Hall J, Viney R, Louviere J, Kenny P, King M. 1999. Discrete choice modelling – how to introduce varicella vaccination into Australia. In Baldry J, ed. Australian Studies in Health Service Administration Nr 87. Sydney: University of New South Wales, School of Health Services Management. p 215–25.

Study 20 Ryan M, McIntosh E, Dean T, Old P. 2000. Trade-offs between location and waiting time in the provision of health care: the case of elective surgery on the Isle of Wight. *J Public Health Med*, 22:202–10.

McIntosh E, Ryan M. 2002. Discrete choice experiments to derive welfare estimates for the provision of elective surgery: implications of discontinuous preferences. *J Econ Psychol*, 23:367–82.

Study 21 Ratcliffe J, Buxton M. 1999. Patients' preferences regarding the process and outcomes of life-saving technology. *Int J Health Technol Assess Health Care*, 15:340–51.

Study 22 Cairns J, van der Pol M. 2000. The estimation of marginal time preference in a UK-wide sample (TEMPUS) project. *Health Technol Assess*, 4:1.

van der Pol M, Cairns J. 2001. Estimating time preference for health using discrete choice experiments. *Soc Sci Med*, 52:1459–70.

Study 23 Gates R, McDaniel C, Braunsberger K. 2000. Modelling consumer health plan choice behavior to improve customer value and health plan market share. *J Bus Res*, 48:247–57.

Study 24 Jan S, Mooney G, Ryan M, Bruggeman K, Alexander K. 2000. The use of conjoint analysis to elicit community preferences in public health research: a case study of hospital services in Australia. *Aust NZ J Public Health*, 24:64–70.

Study 25 Johnson R, Banzhaf M, Desvousges W. 2000. Willingness to pay for improved respiratory and cardiovascular health: a multiple-format, stated preference approach. *Health Econ*, 9:295–317.

Study 26 Morgan A, Shackley P, Pickin M, Brazier J. 2000. Quantifying patient preferences for out-of-hours primary care. *J Health Serv Res Policy*, 5:214–18.

Study 27 Michaels J, Brazier J, Palfreyman S, Shackley P, Slack R. 2000. Cost and outcome implications of the organisation of vascular services. *Health Technol Assess*, 4(11):47–70.

Shackley P, Slack R, Michaels J. 2001. Vascular patients' preferences for local treatment: an application of conjoint analysis. *J Health Serv Res Pract*, 6:152–7.

Study 28 Salkeld G, Ryan M, Short L. 2000. The veil of experience: do consumers prefer what they know best? *Health Econ*, 9:267–70.

Study 29 De Abreu Lourenco R, Hall J, Viney R, Haas M, King M, Kenny P. 2000. Economic evaluation of genetic screening using choice modelling. Paper prepared for the 22nd Australian Health Economists Conference; 2000 Jul 23–24; Surfers Paradise, Queensland. Available from the Centre for Health Economics Research and Evaluation, University of Technology, Sydney.

Study 30 Shanahan M, Gerard K, Louviere J. 2000. Using stated preference discrete choice modelling: examining preferences for breast cancer screening. Paper prepared for the 22nd Australian Health Economists Conference; 2000 Jul 23–24; Surfers Paradise, Queensland. Available from the Centre for Health Economics Research and Evaluation, University of Technology, Sydney.

Gerard K, Shanahan M, Louviere J. 2003. Using stated preference discrete choice modelling to inform health care decision-making: a pilot study of breast cancer screening participation. *Appl Econ*, 35:1–13.

Study 31 Scott A. 2000. Eliciting GPs preferences for pecuniary and non-pecuniary job characteristics. *J Health Econ*, 711:1–19.

Study 32 Godsen T, Bowler I, Sutton M. 2000. How do general practitioners choose their practice? Preferences for practice and job characteristics. *J Health Serv Res Policy*, 5:208–13.

Study 33 Ryan M, Wordsworth S. 2000. Sensitivity of willingness to pay estimates to the level of attributes in discrete choice experiments. *Scott J Polit Econ*, 47:504–24.

Study 34 Roberts T, Bryan S, Heginbotham C, McCallum A. 1999. Public involvement in health care priority setting: an economic perspective. *Health Expect*, 2:235–44.

References

Adamowicz W. 2002. Design and analysis of choice experiments: lessons from environmental economics. Paper presented at Discrete Choice Workshop; 2002 Apr 18–19; Odense. Institute of Public Health, University of Southern Denmark.

Blumenschein K, Johannesson M, Yokoyama K, Freeman P. 2001. Hypothetical versus real willingness to pay in the health care sector. *J Health Econ*, 20:441–57.

Carlsson F, Martinnsson P. 2002. Do hypothetical and actual marginal willingness to pay differ in choice experiments? *J Environ Econ Manag*, 41:179–92.

Clark P. 2002. Testing the convergent validity of the contingent valuation and travel cost methods in valuing the benefits of health care. *Health Econ*, 11:117–27.

De Palma A, Myers G, Papageorgious Y. 1994. Rational choice under an imperfect ability to choose. *Am Econ Rev*, 84:419–40.

Dowie J. 1998. Process utility, conjoint analysis and willingness to pay. Paper presented to the Health Economic Study Group Meeting. 1998 Jul 13–15; University of Galway, Galway, Ireland. In possession of the author.

Emery D, Barron F. 1979. Axiomatic and numerical conjoint measurement: an evaluation of diagnostic efficacy. *Pyschometrika*, 44:195–210.

Fischer G, Hawkins S. 1993. Strategy compatibility, scale compatibility and the prominence effect. *J Exp Psychol: Hum Percept Perform*, 19:580–97.

Gigerenzer G, Todd P. 1999. For The ABC Research Group. Simple heuristics that make us smart. New York: Oxford Univ Pr.

Hadorn D, Hays R, Uebersax J, Hauber T. 1992. Improving task comprehension in the measurement of health state preferences. *J Clin Epidemiol*, 45:233–43.

Hanemann W. 1984. Welfare evaluations in contingent valuation experiments with discrete responses: reply. *Am J Agric Econ*, 69:332–41.

Hanley N, Mourato S, Wright R. 2001. Choice modelling: a superior alternative for environmental valuation? *J Econ Surveys*, 15:435–62.

Hanley N, Wright R, Koop G. 2002. Modelling recreation demand using choice experiments: rock climbing in Scotland. *Environ Res Econ*, 22:449–66.

Heiner R. 1983. The origin of predictable behaviour. *Am Econ Rev*, 73:560–95.

Hensher D, Stopher P, Louviere J. 2001. An exploratory analysis of the effect of numbers of choices sets in designed choice experiments: an airline choice application. *J Air Transport Manag*, 2:373–70.

Huber J, Zwerina K. 1996. The importance of utility balance in efficient choice designs. *J Market Res*, 33:307–17.

Lancaster K. 1966. A new approach to consumer theory. *J Polit Econ*, 74:134–57.

Lloyd A. 2001. The extent of patients' understanding of the risk of treatments. *Q Health Care*, 10(Suppl I):i14–18.

Louviere J. 1988. Conjoint analysis modelling of stated preferences. *J Transport Econ Policy*, 22:93–119.

Louviere J. 2001. Choice experiments: an overview of concepts and issues. In Bennett J, Blamey R, eds. The choice modelling approach to environmental evaluation. Chapter 2. Cheltenham: Edward Elgar. p 34.

Louviere J, Hensher D, Swait J. 2000. Stated preference modelling: theory, methods and applications. Cambridge: Cambridge Univ Pr.

McFadden D. 1973. Conditional logit analysis of qualitative choice behaviour. In Zarembka P, ed. Frontiers in econometrics. New York: Academic Pr. p 105–42.

McIntosh E, Ryan M. 2002. Discrete choice experiments to derive welfare estimates for the provision of elective surgery: implications of discontinuous preferences. *J Econ Psychol*, 23:367–82.

Morrison M, Bennett J, Blamey R, Louviere J. 2002. Choice modelling and tests of benefit transfer. *J Agric Econ*, 84:61–170.

Ohler T, Le A, Louviere J, Swait J. 2000. Attribute range effects in binary response tasks. *Market Lett*, 11:249–60.

Payne JW. 1976. Task complexity and contingent processing in decision making: an information search and protocol analysis. *Org Behav Hum Perform*, 16:366–87.

Payne JW, Bettman JR, Johnson EJ. 1993. The adaptive decision maker. New York: Cambridge Univ Pr. p 34–6.

Pearmain D, Swanson J, Kroes E, Bradley M. 1991. Stated preference techniques: a guide to practice. Hague: Steer Davis Gleave and Hague Consulting Group.

Ratcliffe J. 2000. The use of conjoint analysis to elicit willingness to pay. Proceed with caution? *Int J Technol Assess Health Care*, 16:270–90.

Revelt D, Train K. 1998. Mixed logit with repeated choices: households' choices of appliance efficiency level. *Rev Econ Stat*, 80:647–57.

Ryan M, Gerard K. 2001. Using choice experiments to value health care programmes: current practice and future challenges. Paper presented at International Health Economic Association Conference; 2001 Jul 22–25; University of York, York, UK. Available from Health Economics Research Unit, University of Aberdeen.

Ryan M, McIntosh E, Shackley P. 1998. Using conjoint analysis to assess consumer preferences in primary care: an application to the patient health card. *Health Expect*, 1:117–29.

Ryan M, Scott DA, Reeves C, Bate A, van Teijlingen E, Russell E, Napper M, Robb C. 2001. Eliciting public preferences for health care: a systematic review of techniques. *Health Technol Assess*, 5(5):1–185.

Ryan M, Reeves C, Entwistle V. 2002. Listening to respondents: a think aloud study of discrete choice experiment responses. Paper presented to the Health Economic Study Group Meeting; 2002 Jan 7–9; University of East Anglia, Norwich. Available from Health Economics Research Unit, University of Aberdeen.

San Miguel F. 2000. Testing the assumptions of completeness, stability and rationality of preferences using discrete choice experiments. PhD Thesis. University of Aberdeen.

Schkade D, Payne J. 1994. How people respond to contingent valuation questions: a verbal protocol analysis of willingness to pay for an environmental regulation. *J Environ Econ Manag*, 26:88–109.

Scott A. 2002. Identifying and analysing dominant preferences in discrete choice experiments. *J Econ Psychol*, 23:383–98.

Sen A. 1993. Internal consistency of choice. *Econometrica*, 61:495–521.

Singh J, Cuttler L, Shin M, Silvers J, Neuhauser D. 1996. Medical decision-making for growth hormone therapy using conjoint analysis. *Med Care*, 36:AS31–45.

Smith V. 1993. Non-market valuation of environmental resources: an interpretative appraisal. *Land Econ*, 69:1–26.

Swait J. 2001. A non-compensatory choice model incorporating attribute cut-offs. *Transport Res*, 35:903–28.

Vriens M, Loosschilder G, Rosbergen E, Wittink D. 1998. Verbal versus realistic pictorial representations in conjoint analysis with design attributes. *J Product Innovat*, 15:455–67.

Zwerina K, Huber J, Kuhfield W. 1996. A general method for constructing efficient choice designs. Fuqua School of Business working paper. Durham: Duke University.