International Journal of Technology Assessment in Health Care, 15:3 (1999), 443–457.

A ROLE FOR CONJOINT ANALYSIS IN TECHNOLOGY ASSESSMENT IN HEALTH CARE?

Mandy Ryan
University of Aberdeen

Abstract

The aim of this paper is to demonstrate the use of conjoint analysis (CA) in health services research. Conjoint analysis is first explained, with emphasis on the history of the technique, followed by an explanation of how to carry out such a study and how the results from such a study can be used. The technique is demonstrated with reference to a study that looks at the benefits of in vitro fertilization. It is shown how CA can be used to estimate the relative importance of attributes, the trade-offs individuals make between these attributes, willingness to pay if cost is included as an attribute, and utility or benefit scores for different ways of providing a service. The paper then considers the potential advantages of CA over other, more commonly used benefit assessment instruments. Finally, there is discussion of the issues raised in the design and analysis of CA studies. It is concluded that these issues must be addressed before the technique becomes an established instrument for technology assessment.

Keywords: Benefit assessment, Conjoint analysis, In vitro fertilization

One of the greatest challenges facing health services researchers concerned with technology assessment is the identification and valuation of benefits from health care interventions. Benefit assessment in health economics has been dominated by an assumption that only health outcomes are important. This is evidenced by the large amount of research devoted to valuing health outcomes, using quality-adjusted life-years (QALYs) (16;49;57), and more recently, healthy-years equivalents (26;27;28). Concentration on health outcome fails to allow for the possibility that individuals derive benefit from nonhealth outcomes and process attributes (8;12;30;31;39;42). Nonhealth outcomes refer to sources of benefit, such as the provision of information, reassurance, autonomy and dignity in the provision of care. Process attributes include such aspects of care as waiting time, location of treatment, continuity of care, and staff attitudes. Conjoint analysis has been developed in health economics to take account of factors beyond health outcomes.

I am grateful to all respondents who completed questionnaires and to two anonymous referees for comments on an earlier draft of this paper. Financial support from the Medical Research Council (MRC) and Chief Scientist Office of the Scottish Office Department of Health (SODH) is acknowledged. The views expressed in this paper are those of the author and not MRC or SODH.

https://doi.org/10.1017/S0266462399315312 Published online by Cambridge University Press

The aim of this paper is to demonstrate the potential application of conjoint analysis (CA) in technology assessment. CA is first explained, with the emphasis on the history of the technique, the stages involved in carrying out such a study, and the potential uses of the technique. CA is then applied to consider patient preferences in the provision of in vitro fertilization (IVF). The aim here is to demonstrate the stages involved in carrying out a CA study and the possible uses of data produced by CA studies. Consideration is then given to how CA differs from other methods of benefit assessment and the relative advantages of CA over these techniques. It is concluded that further methodological work in the design and analysis of CA studies is required before the instrument becomes an established technique for technology assessment.

CONJOINT ANALYSIS

CA has its origins in market research (4) and has been widely used in transport economics (2;47;56) and environmental economics (21;34;37;38). The technique was recommended to the U.K. Treasury for examining valuation of quality in the provision of public services (5). It is just beginning to be used in health economics. Its applications in this area are varied and include: looking at optimal service provision; estimating utility or benefit scores for specific studies; estimating willingness to pay (WTP) indirectly; using within the context of randomized controlled trials; looking at patient preferences in the doctor–patient relationship; and looking at prioritizing across clinical service developments.

For example, Ryan and Farrar (42) examined the trade-offs that individuals make between the location of clinic and waiting time in the provision of orthodontic services. The attributes of importance included waiting time, location of first appointment, and location of second appointment. This study presented results on the relative importance of the attributes, the trade-offs that individuals make between these (i.e., how many extra days they were willing to wait for their preferred location), and total utility scores (or benefit or satisfaction) for different ways of providing orthodontic services. Ryan (39) used CA to look at the value of assisted reproductive techniques. This study demonstrated the importance of health outcomes, nonhealth outcomes, and process attributes, as well as the trade-offs that individuals are willing to make between these attributes. It was also demonstrated how CA could be used to estimate WTP indirectly.

Propper (35;36) investigated the monetary value of reducing time spent on National Health Service (NHS) waiting lists. Ryan and Hughes (43) used CA within the context of a randomized controlled trial to identify patient preferences for surgical versus medical treatment of miscarriage. This study demonstrated how the technique could be used to look at the relative importance of attributes identified as different in two arms of a trial. Other similar studies, looking at patient preferences for service provision, include: Bryan et al. (3), who used CA to establish preferences for magnetic resonance imaging in the investigation of knee injuries; van der Pol and Cairns (52), who used CA to investigate patient preferences for blood transfusion support; and Vick and Scott (54), who used CA to assess patient preferences in the doctor–patient relationship. Farrar and Ryan (10) used CA to elicit consultant views for alternative clinical service development. This study went on to estimate cost per unit of benefit ratios for competing clinical service developments.

There are five key stages in the design of a CA study, which are discussed below.

https://doi.org/10.1017/S0266462399315312 Published online by Cambridge University Press

Stage 1: Establishing the Attributes. Stage 1 involves defining the attributes. A number of methods exist to do this, including literature reviews, group discussions, interviews, and direct questioning of individual subjects (39). Alternatively, there may be a predefined policy question. For example, if a health authority is concerned with the trade-offs individuals make between the location of a clinic and the waiting time (i.e., whether to introduce local clinics at the expense of increased waiting time), the attributes will be predefined by this policy question (42). Similarly, if the issue is one of the optimal ways to provide maternity services and possible management plans vary with respect to location, staff involved, continuity of care, and choice involved for women (as in the current debate between obstetrician-led versus midwife-led care), then the attributes are again predefined by the research topic.

Stage 2: Assigning Levels to the Attributes. Levels must be assigned to attributes. The levels must be plausible, actionable, and capable of being traded.

Stage 3: Which Scenarios to Present? In stage 3 individuals are presented with hypothetical scenarios that combine different levels of attributes. The number of possible scenarios increases as the number of attributes and levels increases. Various methods are used to reduce the number of scenarios for inclusion in the questionnaire while still being able to infer utilities for all possible scenarios (1;17;48).

Stage 4: Establishing Preferences. Preferences for scenarios are obtained by surveying patients, service users, and members of the community. The questionnaire design uses one of three methods: ranking, rating, or discrete choices. Economists usually prefer the latter method since it is derived from an area of economics known as random utility theory (13;22). It may also be argued that individuals are used to making such choices since it is the type of exercise in which they engage on a daily basis. Ranking and rating type exercises are seldom carried out by individuals for decision making in the real world. For these reasons the discrete choice approach is developed in this paper. Readers interested in ranking and rating exercises within a CA framework should see Louviere (18).

Within the discrete choice approach, the respondent makes a series of choices. For each he or she chooses (or prefers), the alternative leads to the higher level of utility (or satisfaction or benefit). Thus, the individual would choose health care intervention B over A if:

$$U(A_B, Y, Z) > U(A_A, Y, Z) \qquad \textbf{Equation (1)}$$

where U(.) represents the individual's utility or benefit function; A_B, the attributes of health care intervention B; A_A, the attributes of health care intervention A; Y, the individual's income; and Z, the socio-economic characteristics of the individual that influence his or her utility. In what follows it is assumed that individuals are comparing health care interventions on the basis of health outcomes (HO), non-health outcomes (NHO), and process attributes (P). While the individual knows the nature of his or her utility or benefit function, the researcher will not. This introduces the concept of random utility, where an error term is included in the utility function to reflect the unobservable factors in the individual's utility function. Thus, within the random utility framework, the individual will choose B over A if:

$$V(HO_B, NHO_B, P_B, Y, Z) + \epsilon_B > V(HO_A, NHO_A, P_A, Y, Z) + \epsilon_A \qquad \textbf{Equation (2)}$$

where V(.) is the measurable component of utility estimated empirically; HOj, NHOj, and Pj are the health outcomes, nonhealth outcomes, and process attributes, respectively, of the health care intervention being considered; (j = A, B), Y, and Z are as defined above; and ϵ_j (j = A, B) reflects the unobservable factors in the individual's utility function. Assuming a linear utility function V(.), the utility to be estimated in moving from A to B is:

$$\Delta V = (\beta_B + \Sigma\beta_{iB}HO_{i_B} + \Sigma\beta_{jB}NHO_{j_B} + \Sigma\beta_{kB}P_{k_B} + Y + Z + \epsilon_B)$$
$$-(\beta_A + \Sigma\beta_{iA}HO_{i_A} + \Sigma\beta_{jA}NHO_{j_A} + \Sigma\beta_{kA}P_{k_A} + Y + Z + \epsilon_A)$$

Equation (3)

which can be simplified as:

$$\Delta V = \alpha_c + \Sigma\alpha_i HO_i + \Sigma\alpha_j NHO_j + \Sigma\alpha_k P_k + e$$ **Equation (4)**

where α_c, α_i, α_j, and α_k (i = 1, n,j = n + 1, m, and k = m + 1, l) represent the parameters of the model to be estimated. α_c is the constant term for the model and reflects individual preferences for one type of treatment over another when all attributes included in the model are the same (43).

Stage 5: Analysis of Data. This final stage of a CA study involves using regression techniques to estimate the utility or benefit function specified in Equation 4. Given that multiple observations are obtained from individuals (since individuals are presented with numerous pairwise comparisons), random effects models are usually used to estimate this equation (36). From Equation 4 it is possible to establish:

- *Importance of attributes:* The importance of attributes is indicated by the statistical significance level of parameters α_i, α_j, and α_k. Parameters with a probability level of less than .05 are assumed to be statistically significant and thus important to the respondent.
- *Relative importance of these attributes:* The relative importance of attributes is indicated by their relative size. It is important to be aware of the units of measurement of the attributes when comparing relative size of coefficients.
- *Willingness to trade between attributes:* The ratio of any two parameters in the model shows the marginal rate of substitution (MRS) between these attributes, i.e., the rate at which individuals trade between these attributes.
- *WTP:* WTP is based on the premise that the maximum amount of money an individual is willing to pay for a commodity is an indicator of the utility or satisfaction to her of that commodity. WTP is thus a monetary measure of the value or benefit of a given health care intervention. Traditionally studies have directly asked individuals their WTP, using either the open-ended, payment card, or closed-ended approach (45). If cost is included as an attribute in a CA study, then WTP for individual attributes can be estimated indirectly by looking at the ratio of any attribute coefficient to the cost coefficient.
- *Benefit (or utility) scores:* Utility scores can be estimated for different combinations of levels of attributes.

APPLICATION: USING CONJOINT ANALYSIS TO ASSESS PREFERENCES FOR IN VITRO FERTILIZATION

Methods

Stage 1: Identifying Relevant Effects. Based on literature reviews and previous studies, six attributes were identified as significant predictors of benefit in

Table 1. Attributes and Levels Included in the Conjoint Analysis Study

Attribute	Levels and definition
Process attributes	
Attitude of staff toward you	Bad—uncaring and unsympathetic Good—caring and sympathetic
Continuity of contact with same staff	No—you see many different staff Yes—you see the same staff on all visits
Time on waiting list for IVF attempt	1 month, 3 months, 6 months, 18 months, 36 months
Cost to you of IVF attempt	£0, £750, £1,500, £2,500, £3,000
Health outcome	
Chance of taking home a baby	5%, 10%, 15%, 25%, 35%
Nonhealth outcomes	
Follow-up support	No—no support Yes—you get support

the provision of IVF: chance of taking home a baby; follow-up support; time on the waiting list; continuity of staff; cost; and attitudes of staff (39). These attributes were interesting for a number of reasons. Only one was a health outcome (chance of taking home a baby). There were also nonhealth outcomes (follow-up support) and process-type attributes (waiting time, continuity of staff, and attitudes of staff). The inclusion of cost allowed estimation of the individual's WTP for changes in levels of individual attributes provided in the model as well as overall WTP for IVF clinics with given attributes. Inclusion of the chance of taking home a baby allowed incorporation of risk attitudes into the CA study.

Stage 2: Giving These Effects Levels. Levels for time on the waiting list, cost, and chance of leaving the service with a child were chosen to be representative of the situation at the time in Scotland. Follow-up support was included as a dummy variable, taking on the value of one if it was provided and zero if it was not. Continuity of contact with the same staff took on two possible levels: "yes" and "no", where "yes" was defined as seeing the same staff on all visits and "no" as seeing many different staff. Attitudes of staff toward the individual were also assigned two possible values: "good," defined as caring and sympathetic, and "bad," defined as uncaring and unsympathetic. Table 1 summarizes the attributes and levels included in the CA study.

Stage 3: Which Scenarios to Present. The attributes and levels gave rise to 1,000 possible scenarios ($5^3 \times 2^3 = 1{,}000$). This was reduced to a manageable level of 26 scenarios using the SPSSX ORTHOPLAN procedure (48). The technique results in an orthogonal main effects design. This ensures the absence of multicollinearity and assumes no interactions between attributes.

Stage 4: Valuing Relevant Effects. The discrete choice approach was adopted in this study. The 26 scenarios produced by the statistical design were randomly split into two equal groups, each with 13 scenarios. Within each group one scenario was randomly chosen, and each of the remaining 12 scenarios was compared to this chosen scenario. These two groups formed the basis of two separate CA questionnaires consisting of 12 pairwise choices. Subjects were randomly allocated between these two questionnaires. An example of one of the pairwise choices is shown in Figure 1.

· **Choice 4**	Clinic A	Clinic B
Attitude of staff toward you	Good	Bad
Chance of taking home a baby	25%	35%
Continuity of contact with same staff	No	Yes
Time on waiting list for IVF attempt	18 months	18 month
Cost to you of IVF attempt	£1,500	£3,000
Follow-up support	No	No
	Prefer Clinic A	Prefer Clinic B
Which Clinic would you prefer? (check one box only)	☐	☐

Figure 1. Example of discrete choice question in conjoint analysis questionnaire.

Individuals attending the Assisted Reproductive Unit (ARU) in Aberdeen, Scotland comprised the study sample. In a previous study a satisfaction questionnaire was mailed to all 1,164 individuals who had attended the ARU since it opened in 1989 (39). Among these were individuals who were on the waiting list for IVF, individuals who had had a failed attempt at IVF and were still trying, as well as users who had left the service both with a child and childless. Questionnaires were mailed separately to men and women. The 466 individuals who responded to this satisfaction questionnaire were selected for the CA study. A pilot questionnaire was mailed to a random sample of 52 of these individuals to test whether individuals would complete the CA questionnaire and to see whether trading was taking place among the chosen levels of the attributes. The main questionnaire was mailed to the remaining 414 individuals. Information had previously been collected on respondents' age, sex, whether they had had a child from IVF, whether they were currently undergoing treatment, and how many additional IVF attempts they were willing to have. Respondents were asked to return a slip if they did not want to take part in the study. Two follow-up letters were sent to individuals before they were considered nonrespondents.

Stage 5: Analysis of Data. The equation to be estimated was:

$$\Delta \text{Benefits} = \alpha_1 \text{STAFF} + \alpha_2 \text{CONT} + \alpha_3 \text{WAIT} + \alpha_4 \text{COST} + \alpha_5 \text{CHANCE} + \alpha_6 \text{FOLLOW} + e + u$$

Equation (5)

where Δ Benefits is the change in utility in moving from one IVF clinic A to another B, STAFF is the difference in the staff attitudes between clinic A and B, CONT is the difference in continuity of staff contact, WAIT is the difference in waiting time, COST is the difference in the cost of an attempt at IVF, CHANCE is the difference in the chance of taking home a baby, and FOLLOW is the difference in follow-up support and is the nonhealth outcome. The unobservable error terms are represented by e and u, where e is the error term due to differences among observations and u is the error term due to differences among respondents (36;43). The relative importance of the different attributes is given by α_j ($j = 1,2,3,4,5,6$), and the ratio of the parameters shows the trade-offs between the attributes with α_j/α_4 ($j = 1, 2, 3, 5, 6$) being an estimate of WTP for levels of the individual attributes. The initial utility functions had a constant term that reflects the importance of attributes not included in the model. This was "suppressed" in this study by asking subjects to assume that all aspects of the service, other than those specified in the questionnaire, were identical. Thus, from Equations 3 and 4, $\alpha_c = 0$ (or $\beta_B - \beta_A = 0$)

when there is no difference in the clinics other than those specified for the six attributes included in the model.

RESULTS

Of the 414 questionnaire mailed, 10 were returned because they were sent to the wrong address, 14 were returned by individuals who did not want to take part in the study, 331 questionnaires were returned completed, and there were 59 nonresponses. Of the 331 respondents, 149 were male and 182 female. Forty had a child from IVF, 55 were currently undergoing treatment, and 121 were willing to have an additional attempt. The regression results are shown in Table 2. More detailed statistical analysis of these results found CA to be theoretically valid and internally consistent (39).

Statistical Significance of Effects

Almost all coefficients have the expected sign and are significant at the 1% level, suggesting that these attributes are important in the provision of IVF. The positive signs on attitudes of staff, continuity of care, and chance of taking home a baby indicate that the higher these attributes are in clinic B relative to A, the more likely the individual is to choose clinic B. Similarly, negative signs on cost, time on waiting list, and follow-up support indicate that the lower these attributes are in clinic B relative to A, the more likely the individual is to choose clinic B. The negative sign on follow-up suggests that having this is a disbenefit. This may be explained by the fact that individuals do not want to discuss their infertility.

Size of Effect

A 1% increase in the chance of having a child results in more benefit than a 1-month fall in waiting time or a reduction of one pound in the cost of treatment (as indicated by the size of the respective coefficients). Good staff attitudes are more important than a 6% increase in the chance of taking home a baby (i.e., $0.526 > [0.080 \times 6]$), and continuity of care is slightly more important than a 2% increase in the chance of taking home a baby ($0.180 > [0.080 \times 2]$).

Trade-offs Between Effects and WTP

An individual would be willing to wait almost 12 months longer on the waiting list to have good staff attitudes (0.492/0.042) or to have a reduction in their chance of leaving the service childless of 6% (0.492/0.077). WTP estimates for the various attributes are given by the ratio of the attribute coefficient to the cost coefficient. Individuals are willing to pay: £1,230 for good staff attitudes (0.492/0.0004), £478 for continuity of contact with same staff (0.191/0.0004), £105 to reduce waiting time by 1 month (0.042/0.0004), £193 for an increase in the chance of taking home a child (0.077/0.0004), and £468 extra if the service does not provide follow-up support.

Estimating Benefit Scores and WTP for Changes in Service Provision

Some simulations of different program configurations were conducted to see how CA can be used to look at changes in benefit (or utility) scores resulting from changes in the way that the service is provided (Table 2). In the example presented here, the current IVF clinic is assumed to have bad staff attitudes, no continuity of care, costs users £1,500, a waiting time of 6 months, a chance of taking home a baby of 15%, and no follow-up support. The proposed clinic, in contrast, has good staff attitudes, continuity of care, costs users £2,000, has a waiting time of 3 months,

Table 2. Output from Conjoint Analysis Study

Attribute	Coefficient	Marginal WTP (£)	Current	Proposed	Difference	Benefit marginal utility	Benefit marginal WTP
Process attributes							
Attitudes of staff toward you (0 = bad, 1 = good)	0.492[a]	1,230	0	1	1	0.492	1,230
Continuity of contact with same staff (0 = no, 1 = yes)	0.191[a]	478	0	1	1	0.191	478
Cost to you of IVF, £	−0.0004[a]	N/A	1,500	2,000	500	−0.2	N/A
Time on waiting list for IVF, months	−0.042[a]	105	6	3	−3	0.126	315
Health outcome							
Chance of taking home a baby, %	0.077[a]	193	15	20	5	0.385	965
Nonhealth outcomes							
Follow-up support (0 = no, 1 = yes)	−0.187[a]	468	0	0	0	0	0
Total						0.868	2,988

[a] Significant at 1% level.

https://doi.org/10.1017/S0266462399315312 Published online by Cambridge University Press

a chance of taking home a baby of 20%, and no follow-up support. From the difference in the two clinics, and using the estimated coefficients, the marginal benefit is 0.868, with an associated WTP of £2,988.

Use at a Policy Level

Results from a CA study are potentially very useful to policy makers. Many of the factors identified as nonhealth outcomes and process attributes are capable of being directly influenced by policy makers. Within the context of this study, the results could be used both to help in the design of a given infertility clinic with given resources, as well as for deciding how to spend additional money to improve a clinic. Information on trade-offs between attributes will provide guidance on changing the provision of a service (42). The monetary measure of benefits can be directly compared with costs within the framework of a cost–benefit analysis (CBA). WTP for different scenarios could be generated and comparisons made between them (43). In a similar way estimated benefit scores can be directly compared with information on costs and incorporated into a cost-utility framework, where cost per unit of benefit could be estimated (10;42;43).

HOW DOES CA DIFFER FROM OTHER METHODS FOR MEASURING BENEFITS?

Other Methods of Measuring Benefits

Three main methods have been used by economists to measure benefits: standard gamble (SG), time trade-off (TTO), and WTP. As with the discrete choice CA approach, these techniques all involve the concept of sacrifice (since something is only of value if an individual is willing to sacrifice something for it). The SG involves sacrificing certainty when valuing a health outcome. Respondents are presented with a series of choices between a certain health state (B) or a gamble that may result in either a better health outcome (A) than the certain outcome (with a probability, p) or a worse outcome (C) than the certain outcome (with a probability of $1 - p$). The probability of the best outcome is varied until the individual is indifferent between the certain intermediate outcome and the gamble. The point of indifference is the utility or benefit for the certain outcome (24;25). Using the TTO technique, individuals sacrifice time for improved quality of health (50). Individuals are presented with a series of choices between living for a period t in a specified but less than perfect health state (outcome B) and having a healthier life (outcome A) for a time period h where $h < t$. Time h is varied until the respondent is indifferent between the alternatives. The utility given to the less than perfect state is then h/t. WTP is based on the premise that the maximum amount of money an individual is willing to pay (sacrifice) for a commodity is an indicator of the benefit to them of that commodity (29). Maximum WTP is therefore the utility or benefit measure.

Going Beyond Health Outcomes

Both SG and TTO are limited in terms of their suitability for considering nonhealth outcomes and process attributes. For example, it would be unrealistic to ask individuals the probability level at which they are indifferent between a given intermediate scenario, with a given level of cost and waiting time, and a gamble involving a better scenario (reduced waiting time and cost) with probability p and a worse scenario (higher waiting time and cost) with probability $1 - p$. A similar problem

https://doi.org/10.1017/S0266462399315312 Published online by Cambridge University Press

arises with the application of TTO to valuing outcomes beyond health. The idea of giving up life-years, days, hours, or even minutes for attributes such as information, preferred location, or reduced waiting times may appear unrealistic. In theory the use of WTP allows individuals to value all aspects of care that are described to them, i.e., health outcomes, nonhealth outcomes, and process attributes. The technique has been increasingly used by health economists in this way (6;7;8;41;45). However, despite its growing popularity, many individuals object to the technique for a variety of reasons, including political views concerning paying for health care and the fact that WTP is related to ability to pay.

Using CA to Estimate Quality Weights Within the QALY Paradigm

To date, the development of CA within health economics has taken place alongside the debate about the importance of nonhealth outcomes and process attributes. It has been mainly applied to establish the importance of nonhealth outcomes and process attributes in the provision of health care. A future area of research is to look at the applicability of CA, alongside SG and TTO, to establishing utilities for health outcomes. Three studies have done this to date: Maas and Stalpers (20) adopted CA to estimate utilities for laryngeal patients, Verhoef et al. (53) examined the feasibility of using CA to establish utilities for alternative types of treatment for breast cancer patients, and Harwood et al. (14) applied CA to develop utilities for a handicap measurement scale.

There are a number of potential advantages of using CA over SG and TTO (23). An attractive feature of CA is its ability to take account of risk in a potentially less complicated manner than SG-type questions (though this is ultimately an empirical question). SG has been proposed as the gold standard to estimate utilities under uncertainty. However, it is well recognized that individuals find such questions difficult to answer. The TTO technique has been developed to overcome such problems. However, this technique is not rooted in economic theory. CA has the advantage that risk can be incorporated into the decision-making process by including it as an attribute, and that, following this, the type of question posed to the individual more accurately reflects the type of decision that he or she faces every day. It is also rooted in economic theory.

In everyday life individuals often choose between several options. However, they very rarely consider their probability indifference level, nor the number of years at the end of their life that they are willing to give up for a good or service. Thus, a CA question may be argued to resemble more accurately the type of decisions that individuals make on a daily basis.

There is some evidence that the attribute that is traded off, or gambled with, becomes salient; that is, people give more importance to this attribute (20). This has been shown within SG experiments in which individuals are first asked their probability indifference point between a gamble and certain outcome and are then presented with a pairwise comparison between these two scenarios (i.e., the certainty equivalent and the gamble to which individuals said they were indifferent). In such choice experiments, the certainty equivalent is usually preferred to the gamble (51;55). One possible explanation for this is that the attribute that is traded (risk in this example) is weighted more highly than in a situation where a choice is made. Applying this to the TTO technique, this would show itself by individuals saying they were not willing to trade any number of years for a better quality of life in a TTO question, but choosing a scenario that gave a better quality of life but fewer years to live in a choice question.

DISCUSSION AND CONCLUSION

While CA appears to be a potentially useful instrument, there are a number of issues that must be considered in designing and analyzing a CA study. Some of these issues are unique to CA, while others apply to all methods of eliciting preferences with survey techniques. In stage 1 the researcher must identify important attributes. Exclusion of important attributes will result in the estimated benefit equation being misspecified, thus resulting in inaccurate welfare measures. This is in contrast to the more direct WTP approach, where the researcher does not have to identify the nature of the utility (benefit) function.

After identifying the attributes, levels must then be assigned to them. This raises questions concerning the sensitivity of CA results to the ordering of scenarios, ordering of attributes, and levels chosen. To date there has been little work in this area. The two studies available provide no evidence of ordering effects. For example, the results of an application of CA in a primary care setting provide no evidence that the ordering of scenarios was important (44). A study looking at consultants' views when setting priorities across clinical service developments showed no evidence of framing effects with respect to the ordering of scenarios (10). Further work is needed both to extend this literature and to look at the effect of the attribute levels on the estimated benefit equation.

While including cost as an attribute allows estimation of WTP indirectly, it raises questions concerning the range of such levels and the definition of the cost attribute in a collectively funded health care system. In establishing levels for the cost attribute, future applications of CA should include pilot work to establish the range of monetary values (46). The issues raised here are similar to those raised in devising the bid vector in a closed-ended WTP study. Since IVF is not freely available through the NHS, the cost attribute was defined as "cost to you per IVF attempt." However, this may not be realistic when the health care intervention being valued is collectively funded. Much can be learned here from attempts to apply the direct WTP technique to valuing collectively funded health care interventions. Here, there is a movement away from asking patients about WTP at the point of consumption to the ex-ante insurance-based or taxation-based questions (32;33). Future applications of CA to collectively funded health care interventions should explore this approach.

Stage 3 involves the use of a statistical design to reduce the number of scenarios to a manageable level, while still being able to infer utilities or benefits for all possible scenarios (1;17;48). This raises the potential, particularly if the technique is extended to look at utilities from health outcomes, that some of the scenarios presented by the statistical design may be unrealistic. This in turn raises the question of whether the statistical design should be compromised for realism, or realism compromised for statistical reasons.

A number of issues are raised when eliciting preference using the discrete choice approach. Many of these are common to survey techniques more generally. Data collection for CA has traditionally been undertaken using postal questionnaires. However, there is a trend in the transport economics literature toward using computer software to collect CA data. In choosing which method is most appropriate, the costs and benefits of each need to be compared. The advantage of interviews over a postal questionnaire is that the interviewer can assist in the respondents' understanding and completion of the questionnaire. By introducing computer software, interviewer bias may be reduced. More importantly, the choices

https://doi.org/10.1017/S0266462399315312 Published online by Cambridge University Press

generated for the respondents can be specific to their responses, i.e., the second discrete choice question presented to them would depend on their response to the first.

Any technique that is used to assess benefits should be reliable and valid. Reliability refers to the extent to which the instrument reproduces the same results when the study is repeated within a given time period. Validity is concerned with the extent to which the instrument measures what it is intended to measure. Three types of validity have been identified in the literature: content validity, criterion validity, and construct validity (11;39). Content validity refers to the extent to which a measure takes account of all things deemed important. Criterion validity (sometimes called external validity) is concerned with whether the technique adopted measures what the researcher intends, i.e., whether individuals' stated preferences correspond to their actual preferences. There are two types of construct validity: convergent validity, which measures the extent to which the results are consistent with other measures that are held to measure the same construct, and theoretical validity (often called internal validity), which assesses the extent to which the results are consistent with economic theory, or sometimes more generally, a priori expectations. Content validity is obviously important when conducting a CA study and is directly related to stage 1 of a study (i.e., have all attributes been identified). Within health economics the theoretical validity of CA has been demonstrated (3;35;36;39;43;52;54). One study has established the convergent validity of WTP and conjoint analysis, although further work is obviously needed in this area (40).

There is no work to date that looks at the reliability or external validity of CA within health care. While the former will be relatively straightforward to test for, the latter will be more difficult in health care. It is crucial that future work address the external validity of CA. Researchers must attempt set-up studies to test such validity. Such a test might be to use CA to define those who want to use Clinic A and those who want to use Clinic B, then open both and see who uses one or the other. The question would be whether CA correctly predicted behavior, with actual behavior being the gold standard. However, this test may prove difficult to set up in practice. Alternatively, the researcher could use revealed preference data and compare them with stated preference data from a CA study. A possible test could be set up within a randomized controlled trial (RCT). When conducting RCTs, individuals would be asked whether they have any preference for either arm of the trial. If they do, they are allocated to that arm. If they are indifferent, they are randomly allocated to a trial arm. Thus, at this point of allocation their preferences are elicited. Within a CA study, which may form part of the evaluation alongside the RCT, the same respondent could be asked the same choice, with possible responses being "prefer A" (trial arm), "prefer B" (control arm), or "indifferent." Responses in the CA study could be compared with actual responses.

In the analysis of CA data, assumptions have to be made about the functional form of the utility function. The simplest and most commonly used model is the linear additive model. Research has shown that alternative models seldom result in a significantly better fit than the linear additive model (9). However, further research is needed in the health economics literature before such an assumption can be made.

From the regression equation it is possible to estimate utility scores for different ways of providing a service. Such benefit scores can then be directly compared with costs. However, for this to have meaning at a policy level, it is important to investigate whether the utility scores produced in a CA study are cardinal. For example, if a particular health care intervention has a utility score of 3, and another of 6, can we conclude that the latter is twice as good as the former? This is particularly important when comparing utilities with cost. If a health care intervention costs twice as much, then for efficient decisions to be made we must address the question of whether benefits are at least twice as much. CA has its origins in mathematical psychology, where it has been proved that cardinal data can be obtained from ordinal data (if preferences obey certain axioms) (15;19). These proofs apply to data obtained from ranking and rating exercises. Within the context of the discrete choice approach, research is required on the axioms that underlie random utility theory, whether individuals obey these axioms, and whether the data produced from such studies can be treated as cardinal.

In conclusion, CA is potentially a very useful tool for technology assessment in health care. It can be used to establish the relative importance of different attributes in the provision of care, how these effects are traded, WTP if cost is included as an attribute, and benefit (or utility) scores for different ways of providing a service. It also has the potential to look at the benefits of health outcomes, nonhealth outcomes, and process attributes, as well as how individuals trade between these attributes. The techniques can also potentially be used alongside SG, TTO (within the QALY paradigm), and WTP. However, as with the development of any new instrument, methodological work is needed before CA becomes established alongside other tools for technology assessment.

REFERENCES

1. Bradley, M. *User's manual for the speed version 2.1 stated preference experimenter editor and designer.* Hague Consulting Group: The Hague, 1991.
2. Bradley, M., & Bovy, P. Functional measurement of route choice behaviour: A study of commuting cyclists in the Netherlands. *Behavioural Research for Transport Policy*, 1986, 13, 289–305.
3. Bryan, S., Buxton, M., Sheldon, R., & Grant, A. Magnetic resonance imaging for the investigation of knee injuries: An investigation of preference. *Health Economics*, 1988, 7, 595–604.
4. Cattin, P., & Wittink, D. Commercial use of conjoint analysis: a survey. *Journal of Marketing,* 1982, 46, 44-53.
5. Cave, M., Burningham, D., & Buxton, M. *The valuation of changes in quality in the public services.* London: HMSO, 1993.
6. Donaldson, C., & Shackley, P. Does process utility exist? A case study of willingness to pay for laparoscopic cholecystectomy. *Social Science and Medicine*, 1997, 44, 699–707.
7. Donaldson, C., Shackley, P. & Abdalla, M. Using willingness to pay to value close substitutes: Carrier screening for cystic fibrosis revisited. *Health Economics*, 1997, 6, 145–60.
8. Donaldson, C., Shackley, P., Abdalla, M. & Miedzybrodzka, Z. Willingness to pay for antenatal carrier screening for cystic fibrosis. *Health Economics*, 1995, 4, 439–52.
9. Emery, D. R., & Barron, F.H. Axiomatic and numerical conjoint measurement: An evaluation of diagnostic efficacy, *Psychometrika*, 1979, 44, 195–210.
10. Farrar, S., & Ryan, M. Response-ordering effects: A methodological issue in conjoint analysis. *Health Economic Letters*, 1999, 8, 75–80.
11. Froberg. D., & Kane. R. Methodology for measuring health-state preferences, II: Scaling methods. *Journal of Clinical Epidemiology*, 1989, 42, 459–71.
12. Gerard, K., Turnbull, D., Lange M., & Mooney, G. *Economic evaluation of mammography screening: Information, reassurance and anxiety.* Health Economics Research Unit Discussion Paper No. 01/92. Aberdeen: University of Aberdeen, 1992.

13. Hanemann, W. Welfare evaluations in contingent valuation experiments with discrete responses: Reply. *American Journal of Agricultural Economics*, 1984, 69, 332–41.
14. Harwood, R., Rogers, A., Dickinson, E., & Ebrahim, S. Measuring handicap: The London handicap scale, a new outcome measure for chronic disease. *Quality in Health Care*, 1994, 3, 11–16.
15. Johnson, R. *Beyond conjoint measurement: A method of pairwise trade-off analysis.* Proceedings, Association for Consumer Research, 1975, 353–58.
16. Kind, P. *The design and construction of quality of life measures.* University of York Discussion Paper No. 43. York: University of York, 1988.
17. Kocur, G., Alder, T., Hyman, W., & Aunet, B. *Guide to forecasting travel demand with direct utility assessment*, Washington, DC: U.S. Department of Transportation, 1982.
18. Louviere, J. Analyzing decision making: Metric conjoint analysis. Beverly Hills, CA: Sage Publications, 1988.
19. Luce, D., & Tukey, J. Simultaneous conjoint measurement: A new type of fundamental measurement. *Journal of Mathematical Psychology*, 1964, 1, 1–27.
20. Maas, A., & Stalpers, L. Assessing utilities by means of conjoint measurement: An application in medical decision analysis. *Medical Decision Making*, 1992, 12, 288–97.
21. Magat, W. A., Viscusi, W. K., & Huber, J. Paired comparison and contingent valuation approaches to morbidity risk valuation. *Journal of Environmental Economics and Management*, 1988, 15, 395–411.
22. McFadden, D. *Conditional logit analysis of qualitative choice behavior.* Berkeley, CA: University of California, Berkeley, 1973.
23. McIntosh, E., & Ryan, M. *The wisdom of using conjoint analysis to estimate utility scores in dental health: The case of the third molars.* Paper presented at the Health Economists Study Group Meeting. Galway: The National University of Ireland, 1998.
24. McNeil, B., Weichselbaum, R., & Pauker, S. Fallacy of the five year survival in lung cancer. *New England Journal of Medicine*, 1978, 299, 1397–401.
25. McNeil, B., Weichselbaum, R., Stephen, G. & Pauker, G. Speech and survival. *New England Journal of Medicine*, 1981, 305, 982–87.
26. Mehrez, A., & Gafni, A. Quality-adjusted life years, utility theory, and health-years equivalents. *Medical Decision Making*, 1989, 9, 142–49.
27. Mehrez, A., & Gafni, A. The healthy-years equivalents: How to measure them using the standard gamble approach. *Medical Decision Making*, 1991, 11, 140–46.
28. Mehrez, A., & Gafni, A. Healthy-years equivalents versus quality-adjusted life years. *Medical Decision Making*, 1993, 13, 287–92.
29. Mitchell, R. C., & Carson, R. T. *Using surveys to value public goods.* Washington, DC: Resources for the Future, 1989.
30. Mooney, G. What else do we want from our health services? *Social Science and Medicine*, 1994, 39, 151–54.
31. Mooney, G., & Lange, M. Ante-natal screening: What constitutes benefit? *Social Science and Medicine*, 1993, 37, 873–78.
32. O'Brien, B., & Gafni, A. When do the 'dollars' make sense? Towards a conceptual framework for contingent valuation studies in health care. *Medical Decision Making*, 1996, 16, 288–99.
33. Olsen, J. A., & Donaldson, C. Helicopters, hearts and hips: Using willingness to pay to set priorities for public sector health care programme. *Social Science and Medicine*, 1998, 46, 1–12.
34. Opaluch, J., Swallow, S., Weaver, T., Wessells, C., & Wichelns, D. Evaluating impacts from noxious facilities: Including public preferences in current siting mechanisms. *Journal of Environmental Economics and Management*, 1993, 24, 41–59.
35. Propper, C. Contingent valuation of time spent on NHS waiting list. *Economic Journal*, 1991, 100, 193–99.
36. Propper, C. The disutility of time spent on the United Kingdom's National Health Service waiting lists. *Journal of Human Resources*, 1995, 30, 677–700.

37. Rae, D. *Visibility impairment at Mesa Verde National Park: An analysis of benefits and costs of controlling emissions in the Four Corners area.* Report prepared for the Electric Power Research Institute. Boston: Charles River Associates, 1981a.
38. Rae, D. *Benefits of improving visibility at Great Smoky National Park.* Draft report prepared for Electric Power Research Institute. Boston: Charles River Associates, December, 1981b.
39. Ryan, M. *Economics and the patient's utility function: An application to assisted reproductive techniques.* Unpublished Ph.D. thesis. Aberdeen: University of Aberdeen, 1995.
40. Ryan, M. *Establishing the convergent validity of willingness to pay and conjoint analysis for eliciting preferences.* Paper presented to the Health Economists' Study Group Meeting, University of Brunel, July 1996.
41. Ryan, M. Using willingness to pay to assess the benefits of assisted reproductive techniques. *Health Economics*, 1997, 5, 543–58.
42. Ryan, M., & Farrar, S. *A pilot study using conjoint analysis to establish the views of users in the provision of orthodontic services in Grampian.* Health Economics Research Unit Discussion Paper No. 07/94. Aberdeen: University of Aberdeen, 1994.
43. Ryan, M., & Hughes, J. Using conjoint analysis to value surgical versus medical management of miscarriage. *Health Economics*, 1997, 6, 261–73.
44. Ryan, M., McIntosh, E., & Shackley, P. Methodological issues in the application of conjoint analysis in health care. *Health Economics Letters*, 1998, 7, 373–78.
45. Ryan, M., Ratcliffe, J., & Tucker, J. Using willingness to pay to value alternative models of antenatal care. *Social Science and Medicine*, 1997, 44, 371–80.
46. San Miguel, F., Ryan, M., & Mcintosh, E. Applying conjoint analysis in economic evaluations: An application to menorrhagia. *Applied Economics*, 1999, in press.
47. Sheldon, R., & Steer, J. *The use of conjoint analysis in transport research.* PTRC Summer Annual Meeting, Warwick, 1982.
48. SPSS Incorporated. *Statistical package for the social sciences.* Chicago: SPSS, 1989.
49. Torrance, G. Measurement of health state utilities for economic appraisal: A review. *Journal of Health Economics*, 1986, 5, 1–30.
50. Torrance, G., Thomas, W., & Sackett, D. A utility maximisation model for evaluation of health care programs. *Health Services Research*, 1972, 7, 118–33.
51. Tversky, A., Sattath, S., & Slovic, P. Contingent weighting in the judgement of choice. *Psychological Review*, 1988, 95, 371–84.
52. van der Pol, M., & Cairns, J. Establishing preferences for blood transfusion support: An application of conjoint analysis. *Journal of Health Services Research and Management*, 1998, 3, 70–76.
53. Verhoef, C., Maas, A., Stalpers, L., et al. The feasibility of additive conjoint measurement in measuring utilities in breast cancer patients. *Health Policy*, 1991, 17, 39–50.
54. Vick, S., & Scott, A. What makes a perfect agent? A pilot study of patients' preferences in the doctor-patient relationship. *Journal of Health Economics*, 1998, 17, 586–606.
55. Von Winterfeldt, D. Additivity and expected utility in risky multiattribute preferences. *Journal of Mathematical Psychology*, 1980, 21, 66–82.
56. Wardman, M. A comparison of revealed preference and stated preference models. *Journal of Transport Economics and Policy*, 1988, 22, 71–91.
57. Williams, A. Economics of coronary artery bypass grafting. *British Medical Journal*, 1985, 1, 326–29.