ARTICLE

Stated preference methods in health care evaluation: an emerging methodological paradigm in health economics

John FP Bridges

Department of Epidemiology and Biostatistics, Health Services Research Division, Case School of Medicine, Cleveland, OH, USA

Abstract: One focus of health economics is the trade-off between limited resources and the (health) needs of a community. Cost-effectiveness analysis (CEA), while being one of the most accepted evaluation methodologies in health economics, does not account for many important costs and benefits of health care interventions. Some health economists have attempted to modify CEA to account for these deficiencies, while others have been working on alternative methodologies. One group of alternative methodologies can be described as stated preference techniques. These aim to measure both health and non-health outcomes (ie costs and benefits), and include qualitative analysis, conjoint analysis (often referred to as discrete choice analysis/modelling) and willingness to pay (or contingent valuation). This paper provides an overview of stated preference techniques in health economics, with particular focus on their strengths as compared with traditional evaluation methods in health care. The limitations and policy implications of these methods are also discussed.

Keywords: stated preferences, qualitative analysis, conjoint analysis, willingness to pay

Introduction

Economic evaluations of health care interventions are performed in order to inform and manage the trade-offs between scarce resources and the (potentially unlimited) need for health care services (Graff Zivin and Bridges 2002, p 135). The most prevalent tool used in this endeavour is cost-effectiveness analysis (CEA) (Elixhauser et al 1998, p MS4–MS5), which has become one of the cornerstones of health technology assessment and pharmacoeconomics. One of the limitations associated with CEA, however, is the inability to include all the health and non-health outcomes (both costs and benefits) faced in making health-related choices (Mooney 1994; Lindholm et al 1998, p 808; Ryan 1999, p 535). This, plus a number of more technical issues, has prompted many authors in health economics to question the use of CEA (Birch and Gafni 1992). Others have worked to either substantially refine CEA (Stinnett and Mullahy 1998) or replace it with an alternative method (Graff Zivin and Bridges 2002, p 138).

One group of methods that have been proposed as a substitute for CEA are known as stated preference techniques. These methods aim at valuing the costs and benefits of a health intervention from the perspective of the individual or society. These methods include qualitative analysis (Coast 1999, p 345–6), conjoint analysis (sometimes referred to as stated preference discrete choice analysis/modelling) (Ryan 1999, p 536–7) and willingness to pay (often referred to as contingent valuation) (Donaldson 2001, p 181–2). Despite their growing popularity among health economists, stated preference techniques are not well known among policy makers and physicians, and, to date, have rarely been used in health technology assessment or pharmacoeconomics.

This paper provides an overview of stated preference methods and demonstrates their benefits over more traditional evaluation techniques in health care.[1] Emphasis is given to the pitfalls of the dominant paradigm of economic evaluation of health care, extra-welfarism, and the benefits of using an alternative paradigm referred to as welfarism. Three types of stated preference techniques: qualitative analysis, conjoint analysis and willingness to pay (WTP) are outlined. Some of the potential limitations of stated preference techniques are presented, as are various methods

Correspondence: John FP Bridges, Department of Epidemiology and Biostatistics, Health Services Research Division, Case School of Medicine, 10900 Euclid Avenue, Cleveland, OH 44106-4945, USA; tel +1 216 368 6962; fax +1 216 368 3970; email healtheconomics@hotmail.com

that have been suggested to overcome the limitations. Finally, important implications for policy makers raised by stated preference techniques are discussed.

The benefits of stated preferences over cost-effectiveness analysis

While the optimal research methods for any evaluation will depend on the context, focus and perspective of the research question, some core elements of an economic study are necessary to make relevant and grounded policy conclusions. It can be argued that stated preference techniques are preferable to traditional CEA evaluations as they are designed to be rooted in the experience and preferences of the individual or community. There are at least five reasons why stated preferences are preferable to CEA. Specifically:

1. While CEA techniques have been extended to include many different aspects of health care interventions, such as quality of life and equity, they are focused on outcomes which are aggregated into a single measure. Stated preference methods are more holistic and can focus on all relevant aspects of the health intervention, including the process of care (Gerard and Mooney 1993; Mooney and Lange 1993; Donaldson and Shackley 1997, p 699–701; Ryan et al 1999, p 537–8). Being broadly based, they are more relevant to the people involved in the intervention (patients, physicians and policy makers).
2. CEA can only be used to rank interventions on predetermined criteria and cannot be used to improve the interventions under investigation. Stated preference techniques, however, focus on the individual attributes of the intervention/s under investigation, and thus can be used to compare interventions, identify improvements and suggest new interventions. This is achieved by not only examining the total worth of an intervention, but also the value (or what is often referred to as part worth) of each of the attributes of an intervention.
3. One of the problems associated with CEA is its reliance on expected utility theory (Kahneman and Tversky 1982). While a number of authors have written on the importance of risk, both from the perspective of the social planner and the individual (Graff Zivin and Bridges 2002, p 136–7), there is still no clear consensus on how to incorporate risk preference into CEAs. In contrast, stated preference techniques do not require assumptions about risk preferences but rather utilise individuals' or society's views on risk.
4. While there is still much debate surrounding the theoretical foundations of CEA (Birch and Gafni 1992, 1994), stated preference techniques are well supported by economic theory (Ryan 1999, p 536–7; Donaldson 2001, p 184). This makes stated preference techniques more conceptually appealing, and easier to explain to policy makers and physicians. CEA techniques are harder to explain, as outcomes measures may be based on population-based metrics or implicit concepts such as quality adjusted life years (QALY) (Hoffmann and Graf von der Schulenberg 2000, p 185).
5. Stated preference evaluation methods are built upon a welfarist perspective (one that focuses on consumer or social welfare) rather than the more traditional extra-welfarist perspective used in many analyses (Birch and Donaldson 2003, p 1123). Stated preferences are preferable because they place the patients'/consumers' preferences at the centre of the analysis. Thus, these methods focus on increasing individual or social welfare, rather than a predetermined objective.

Welfarist versus extra-welfarist

To expand on the fifth reason, extra-welfarism is the dominant paradigm in economic evaluation of health care. It focuses not on the maximisation of individual (or social) welfare, but rather on some defined 'goal' of the health sector. This goal is normally defined by the maximisation of some health outcome, or adjusted health outcome using quality of life or equity weights (Cuyler 1990). An objective function in CEA may be used to measure the efficiency of proposed interventions, but it lacks transparency and overlooks many of the complexities of social decision making. There is no guarantee, for example, that the objective reflects an individual's or society's preferences in relation to all (or even any) health care decisions. Extra-welfarism is a normative science (Blaug 1998, p S63–4), involving assumptions concerning the choice of health outcome and the population valuing the health outcome, and it has been at the centre of much criticism in recent years (Mooney 1998, p 1171–3; Birch and Donaldson 2003, p 1123–5). Welfarism is grounded in economic theory and aims at maximising the wellbeing of the individual or society.

From a pure welfarist perspective, cost-benefit analysis (CBA) would be the optimal form of evaluation (Johannesson and Jonsson 1991); however, such analyses are very difficult to perform (Robinson 1993). Nevertheless, traditional methods of CEA are not an adequate substitute

(McIntosh et al 1999, p 358). The stated preference techniques discussed in this paper focus on preferences in a multidimensional framework that is flexible enough to accommodate many evaluation scenarios. As such, these methods offer a way of valuing resources without a need to perform a full blown CBA. One limitation of the stated preference framework, as compared with CBA, is that it does not necessarily combine production data (ie costs) into the analysis, but places all emphasis on the valuation.

Qualitative analysis

Qualitative methods are rarely used in economics, except for case studies in relation to management. This is despite the obvious role that qualitative analysis could play in the formation of axioms prior to theoretical and empirical modelling. Probably due to the interdisciplinary nature of health economics, qualitative methods are more frequently used there than in economics in general (Coast 1999, p 345–6). Nevertheless, their use is still very limited (for example, see Coast 2001; Litva et al 2002), and there are significant hurdles to publication due to lack of familiarity and understanding of the techniques.

Qualitative methods

Qualitative methods aim at understanding how individuals or groups perceive and/or operate in a particular environment. They use direct observation, open-ended interviews and focus groups to draw empirical observations (Patton 1987, p 12). Rather than being used to test a theory or hypothesis, qualitative methods are used to gather evidence for existing theories, formulate new theories, or derive hypotheses (Patton 1987, p 7–10; Coast 1999, p 347). In this way, qualitative research assists the researcher and respondents to explore causation and interaction to explain personal and social behaviour (Patton 1990, p 13–14). Theory is developed inductively, where the experiences of the study population (obtained by observation and/or interview) are the basis of the conclusions.

In qualitative analysis the researchers are the instrument, ie the means by which data will be collected (Guba and Lincoln 1981, p 113; Sorrell and Redmond 1995), and, as such, they have more responsibility for the conclusions than researchers undertaking quantitative analyses. Hence, the qualitative researcher may be the cause of biases in the results; however, many of the potential 'researcher' biases can be moderated by the use of appropriate techniques. These include training, formal record keeping and triangulation (Patton 1987, p 60–5). Triangulation involves systematic variation of: (1) the source of data; (2) the investigator; (3) the perspective/paradigm/theoretical slant of the research; and/or (4) the method of analysis, in order to converge at the study conclusions from divergent starting points or processes.

In recent years, computer programs have been developed to assist in the analysis of qualitative data – especially data retrieved from interviews and focus groups (St John and Johnson 2000, p 393). Common programs such as NUDIST, ATLAS.ti, and Folio Views® can be used for:

1. Note taking in the field and editing notes
2. Transcribing audio recordings into text
3. Text storage and searching text
4. Memo writing and commentary on the text
5. Coding and attaching keywords to the text
6. Content analysis, frequency analysis, theory building and hypothesis testing.

Some claim that this software is beneficial in focusing the qualitative researcher on the content of data rather than on the methodology (Richards and Richards 1994). Others have found that the restrictions imposed by computer programs have negative consequences, as they make the analysis less realistic and natural (Seidel 1991). The decision to use computer-assisted analysis should be based on the context and aim of the analysis. Furthermore, as previously discussed, it may be beneficial to use several methods of analysis so as to highlight any sources of bias.

Interpreting the results

One of the key components of a qualitative study is known as constructivism or naturalism. This refers to the non-contrived setting and tone of the interviews and analysis (Guba and Lincoln 1981, p 113). While this may reduce researcher or study bias, it may suggest some theories that were unintended by the initial research objectives. Such unsolicited responses may be beneficial to general understanding, but they may be inconsistent with the limited bounds of traditional 'rational' economic analysis (Coast 1999, p 349). On the other hand, 'orthodox' economics has been criticised for being too limited in scope, especially by feminist researchers and by those advocating the use of normative economics. Thus, the incorporation of formal qualitative analysis into the health economist's 'tool box' may be beneficial in broadening perspectives.

The analysis of qualitative data can take at least two different perspectives on the experience and perceptions of the respondents/subjects. First, 'relativism' holds that there

is no single reality and that each individual's experience may be different. Alternatively, 'realism' assumes that there is a single knowable reality, and is more consistent with traditional economic theory (Lawson 1995, p 13). As discussed below, stated preference methods can be used to understand differences across individuals or to aggregate/average across a population of individuals. They present a departure from the more strict realism of CEA and may be used to change the focus of analysis of health evaluation from populations to subgroups and/or individuals (Sculpher and Gafni 2001, p 317).

Qualitative methods are particularly beneficial when they are used in conjunction with quantitative methods (often referred to as mixed methods), or when used in the development of quantitative instruments (Louviere et al 2000, p 257–8). Furthermore, qualitative data can be used at the end of a quantitative analysis as a method of understanding the quantitative results more fully (Coast 1999, p 351). Qualitative methods may also be beneficial on their own, but, given the dominance of quantitative methods in the economic and medical literature, mixed methods may be more acceptable to a potential reader. More needs to be done to educate the users and providers of economic and policy analysis of the relative benefits of pure qualitative methods.

Conjoint analysis

Conjoint analysis attempts to measure preferences by asking respondents to choose among stylised scenarios (clinical vignettes) in order to inform clinical decisions and/or policy. The much talked about utility function is merely a numerical representation of preferences. The concept of preferences underpins almost all consumer and social choice theory, and is quite basic. If an individual is offered two options, then the individual must either prefer one over the other, or must prefer both options equally (a concept called indifference). This holds for all such options, whether they are feasible (affordable) or not. Medical decision making fits this simple model well, as a high proportion of medical decisions involve binary choices. For example, a patient may or may not seek treatment, or a physician may or may not refer, or may recommend one treatment over another, or may order one type of test over another. In decision making, revealed preferences (evidence of the actual choices made) are often complicated by constraints or selection mechanisms which limit inference to other settings (Bridges 2003, p 2543–4). Stated preferences allow us to model choices under many different scenarios, under different constraints and across any population. As such, they offer a flexible mechanism for not only evaluating current programmes, but also new, future, or even potential (ie hypothetical) programmes.

Conjoint analysis methods

Conjoint analysis has been the focus of much recent interest in health economics (Ryan 1999, p 536–7; Ryan and Gerard 2003, p 55–64) and recent acclaim in economics, with McFadden winning the Nobel Prize in Economics in 2000 for his seminal paper on the topic (McFadden 1974). The technique has been shown to be a valid predictor of real world decisions (Louviere et al 1981, p 42). There are a number of authors that have written in-depth guides for conjoint analysis (see Louviere et al 2000, p 255 for a seven step method). The methods are relatively simple and flexible enough to meet many different scenarios.

Beyond forced binary choice methods, many other forms of conjoint analysis have been discussed (see Orme 1996, p 2–7), and are used in areas such as marketing and product development. These include the use of rating scales to map the relative benefits of one scenario over another, presenting respondents with more than two options and/or allowing the respondent to not accept any of the options presented. While the alternative techniques offer some benefits, we will limit our discussion to forced choice paired scenarios, as this method has been most widely used in health economics and has greater theoretical support (McFadden 1974).

Planning a conjoint analysis

Unlike CEA, conjoint analyses are extremely flexible. However, this does not eliminate the need for careful planning, as with all stated preference techniques, before one begins. Specifically, four important questions must be addressed:

1. *What is the domain of the analysis?* Is the analysis aimed at understanding: a single type of intervention, such as screening; a certain disease, such as asthma; or a certain type of service, such as community health centre?
2. *What is the population of interest?* Is the population of interest the direct recipients of an intervention, their advocates (eg primary care providers), the general population or some other decision maker/s?
3. *Why are you interested in understanding preferences?* Is the analysis driven by an interest in determining the

optimal treatment, modifying an existing treatment, or trying to understand the level of need in a community?

4. *How do decisions normally get made regarding the intervention under investigation?* What type of information is normally needed to make a decision, and what are the 'normal' circumstances of decision making (eg is a decision made immediately, and are family members involved)? These aspects are important if the analyst is to develop conjoint tasks that resemble the normal setting for decision making.

Gaining an understanding of the problem

As stated above, qualitative analysis can be used as a stand-alone tool or as a means of improving quantitative analysis. While resource constraints may not permit a full qualitative analysis before every conjoint analysis, such information can be gained through other sources, including literature reviews, expert advice and previous experience. The aim is to identify which attributes are important in making a decision, and what ranges of attributes (ie attribute levels) are plausible for decision makers (Louviere 1988, p 257–8). Ryan (1999, p 543–4) has demonstrated that conjoint analysis can include a broad selection of attributes including health outcomes, non-health outcomes and process issues.

Developing and piloting the conjoint tasks

Once the domain of the conjoint analysis is chosen, including the attributes and attribute levels, attention can then be given to the tasks that will be presented to the respondents. This involves both the development and the piloting of the tasks. Note that these are not two distinct stages, but rather they comprise an iterative process, which involves four main components:

1. *The selection of the scenarios to be presented to the respondents.* Specifically, this involves the fine-tuning of the attributes and attribute levels that will constitute the scenarios, and the combinations and order in which the various scenarios will be presented to the respondents. Given that it may be impossible (and unnecessary) to include all combinations of the chosen attributes and attribute levels, a process referred to as 'experimental design' is undertaken to limit the scenarios. Care must be taken when limiting the possible attribute combinations, a process referred to as fractional design, as a poor design will hinder subsequent estimation (Louviere 1988, p 35). One method used to select a subset of attribute combinations is referred to as orthogonal factor design (Louviere 1988, p 43–7).
2. *The presentation of the tasks to be performed by the respondents.* One has to decide on how best to convey the attributes to the individual. This may require the use of particular language revealed in the qualitative analysis or on graphical/visual representations of the attributes and attribute levels.
3. *The piloting of the tasks on potential subjects and review of results*. The validity of a task can be verified through standard techniques. A particular question could be repeated to check consistency of respondent choice. In addition, or instead, it could be verified that the respondents made the choices that they wanted to make, and that all the appropriate attributes were presented or addressed.
4. *Refinement and implementation of modifications to the individual tasks to arrive at the final design and stated preference instrument.* This entire process is an iterative one, but the aim of the refinements is to develop an instrument that is relevant, valid and visually appealing to the respondent.

Power calculation

Given the semi-qualitative nature of conjoint analysis, there are no hypotheses that allow a formal power calculation (Orme 1998, p 1). Louviere et al (2000, p 261–5) offer one method to get an approximate power calculation, with a statement of hypotheses that is based on the standard error of proportions known as the 'exogenously stratified random samples'. One of the limitations of their calculation is that it does not account for the complexity of the tasks presented to the individual, ie the actual number of attributes and attribute levels. As one adds more attributes or (categorical) levels, the model will have more parameters to be estimated. There have been a number of rules of thumb that have been suggested, but these are non-statistical (Orme 1998, p 6–9). Technically, the minimum sample size for a conjoint model is one, given that the respondent answers more tasks than parameters to be estimated in the model (Orme 1998, p 6). In this case, and in others where the entire population participate in the conjoint task, sampling theory does not apply, and a power calculation based on sampling theory is inappropriate.

Performing the task and estimating the model

There are a number of different methods for the administration of conjoint tasks, such as through a paper-based survey, a computer-based survey, or the use of an interviewer. The choice of method should be based not only on the abilities of the respondents, but also on the sensitivity of the interventions being evaluated. Once the tasks have been administered, the data analysis and model estimation can take place.

Before estimating the model, one can analyse the quality of the data through what is known as the rationality test. This focuses on question pairs where one observation clearly dominates the other (ie without any trade-offs needed). Respondents who choose the clearly dominated option may be deemed 'irrational', which typically implies they have not understood the task. One can also look for and delete non-traders, or individuals exhibiting 'lexicographic' preference, ie preferences dominated by a single attribute. When performing such tests, however, one has to be careful not to impose any 'outside' or paternalistic views of rationality on the choices of the individuals.

The model is estimated by regressing scenario choices (the dependent variable) upon the variation in the attributes across the scenarios presented to the respondent(s) (the independent variables). Given the properties of the data collected by a conjoint analysis (ie that they have dichotomous dependent variables, and the data are clustered by respondent), special care has to be taken to analyse them. A number of regression techniques can be used to estimate the data (eg logistic, probit, linear probability models) depending upon the assumption the investigator is willing to make about the distribution of the data. The clustering of the data can be simply handled through the estimation of robust standard errors, or by including variables that account for the variation across respondents.

Presentation and interpretation of the results

The main result of a conjoint task is the estimated value function, ie the results from the regression. The estimated parameters are interpreted as the effect on the likelihood that a scenario would be accepted following a marginal change in that particular attribute. If a change in an attribute is associated with an increase in the likelihood of acceptance, then, under random utility theory, it is assumed that the individual has a positive preference for the change.

In addition to the regression parameters, some researchers are interested in specific trade-offs between estimates, or what an economic theorist would call marginal rates of substitution (MRS) between attributes. If one has estimated a simple linear model with no interaction terms, then the MRS is the (negative) ratio of any two of the estimated parameters (Louviere et al 2000, p 61). If one of the attributes is measured in monetary units, then the MRS between another attribute and the (numeraire/denominator) attribute can be interpreted as a marginal WTP (Louviere et al 2000, p 280). While it is popular to use conjoint analysis results to derive marginal WTP values, this is not a necessary component of conjoint analysis. In fact, there are a number of potential faults with this method of measuring WTP, as has been identified in the literature (Ratcliffe 2000, p 270; Skjoldborg and Gyrd-Hansen 2003, p 479).

The use of a numeraire such as price, and estimation of the marginal WTP for each variable does, however, allow the ranking of the relative importance of the attributes. A number of other alternative methods have been proposed. For example, one might be interested in what percentage of decisions was affected by a particular attribute. Care has to be taken here, as the importance of a particular factor will depend on its attribute levels. The more extreme the levels of the attribute, the more important the attribute will appear to be. For example, if price of the intervention is included as an attribute with two levels, then the range of these values will impact the importance of price compared to the other variables. Thus, any interpretation on the relative importance of attributes can only be considered within the context of the model, and not generalised to all decision making scenarios.

As part of the interpretation of the results, one may wish to examine clusters or subgroups of respondents, and the observed variation in their preferences. To facilitate subgroup analysis, one may include other questions in the survey (in addition to the conjoint tasks) that focus on the respondents' background, beliefs and other possible covariates. Also, open-ended questions may be included about either the interventions under investigation, or the respondents' comfort about their understanding and completion of the conjoint tasks.

Policy simulation is one of the important stages of stated preference models. It is even more advantageous if the model includes both current and possible attributes, allowing potential changes to the status quo to be assessed. Policy simulation can be done by modifying (restricting) the exogenous variables and illustrating graphically (Bridges and Hanson 2001, p 97).

Willingness to pay

WTP methods are often referred to as contingent valuation. They attempt to estimate the value of a medical intervention from the perspective of the individual patient/consumer (Donaldson and Shackley 1997, p 700; Donaldson 2001, p 181), or society (Olsen and Donaldson 1998, p 2–3). The WTP method has been used in many other areas of applied economics to assess the value of commodities and public goods.

The aim of the method is to ascertain the maximum amount of money that a patient would be willing to pay, hypothetically, to receive/consume the commodity/service. This represents a monetary valuation for the change in utility that the consumers experience when they consume/receive the commodity/service. Thus, WTP represents the 'value' of the commodity or service (Donaldson 2001, p 181).

While WTP may seem to be relatively straightforward (ie 'what is your valuation for programme *x*?'), there are a number of important aspects to eliciting a valuation. These relate to the influence on valuation estimates of the way in which the question is formed (van der Pol 2003, p 99). However, there have been a number of advances made in recent years that have led to more consistent estimation of WTP (Donaldson 2001, p 182–5). The WTP valuation task relies upon five important decisions:

1. *One has to decide whether to focus on either personal or social preferences.* This decision will depend on the perspective taken in the analysis; however, there are a number of additional complicating factors. It is unclear whether personal preferences include altruism or are solely based on the personal benefits derived by the individual. For example, an individual may value a type of care that they will never receive because they want the service to be available for others (eg an adult may value paediatric services). It is clear, however, that an individual will have a different stated preference if they are asked to take a social perspective – Mooney (1998, p 1174) refers to this distinction as having *levels of preferences*. Social preferences will contain more altruism. In countries like the USA, it may be hard for a respondent to conceptualise social preferences, while in European countries there is a clearer distinction. The ability of an individual to conceptualise social preferences is linked to the concept of social capital – a concept that relates to there being a sense of community or collectiveness, rather than pure individualism. This concept is best illustrated by Lomas (1998, p 1182) who writes:

   > Put simply, individuals (and their ill-health) cannot be understood solely by looking inside their bodies and brains; one must also look inside their communities, their networks, their workplaces, their families and even the trajectories of their life.

2. *One has to decide upon the amount of background information to be given on the interventions to be valued.* While one might think that more information is better, detailed information may lead to confusion or systematic variations in interpretation that could bias the results. One needs to present a parsimonious description of the intervention. This would reflect the level of information that a physician might give before asking a patient to make a treatment decision. Furthermore, when expressing risks to respondents, it is unclear whether probabilities should be expressed in individual (Gafni 1991, p 1246–52; Morrison and Gyldmark 1992, p 233–43) or population terms (O'Brien and Gafni 1996; Olsen and Donaldson 1998, p 11–12). Population risks can be considered certain (ie five people out of one hundred will be affected) while individual risks invoke some uncertainty (a one in 20 chance) (see Graff Zivin and Bridges 2002, p 136–7). Again, the choice will depend on the focus of the analysis and the extent to which one wants to focus on concepts of risk.
3. *One has to make a decision whether to focus on WTP or willingness to accept (WTA).* WTP seeks to find a money metric measure of benefit to an individual or society arising from receipt or access to an intervention. WTA is a money metric measure of the loss of welfare experienced when an individual or society loses or is excluded from an intervention. Bateman et al (2002, p 24–8) offer an excellent illustration of the links between the concepts of WTP and WTA and those of compensating and equivalent variation that are discussed in welfare economics. While one might think that WTP and WTA might give identical results, there is most often a difference observed between the two measures, dependent on a number of factors (for a full discussion, see Hanemann 1999, p 42–96).
4. *One has to decide if the task will focus on marginal or absolute estimation of WTP.* An absolute WTP measures the total valuation of each of the interventions presented to the respondent and then compares these valuations. Alternatively, a marginal estimation will try to measure the differences in the valuations, without reference to the total valuation of any of the interventions. The

determination of marginal WTP for a preferred treatment is a two-part process. First, one must ask if the individual would prefer the traditional treatment (which may be no treatment) or the new intervention. Then, one asks how much the individual would be willing to pay to have their preferred treatment instead of the other. Determination of absolute valuation for an intervention has several limitations. For example, individuals have varying reference points; thus, prospect theory suggests that there may be inconsistent choices in the aggregate (Stalhammar 1996, p 242–3). As well as being more reliable, the marginal approach is more consistent with economic theory (Donaldson 1999, p 552–4). Marginal WTP is also clearer in eliciting value, rather than cost, as it aims to calculate marginal valuations, rather than overall value, and is less affected by ability to pay (Donaldson 1999, p 554).

5. *One has to decide the method that respondents will use to report their WTP*. There are a number of possible methods identified in the literature (Dong 2003, p 850). To measure WTP, one may: ask a direct question (what is your (relative) valuation for *x*?); use cards with existing dollar amounts on them (and ask the respondent to circle the highest amount they would be willing to pay); or use a series of binary questions (Arrow et al 1993). The last approach makes the valuation process very similar to the conjoint analysis method.

Potential pitfalls of stated preference methods

There are pitfalls associated with stated preference techniques. These include variation in tastes across individuals, inability to comprehend/participate in the associated tasks, inertia and flexibility in preference, ability to pay and nesting of valuations. Therefore care has to be taken in performing these methods and interpreting their results. As discussed below, however, many of these issues are not unique to stated preference techniques, but arise, in one form or another, in nearly all types of economic evaluation in health.

Variation in tastes across subgroups

While stated preference techniques can be used to ascertain both individual and societal-level preferences, it is uncertain whether one should survey specific patient subgroups (Sculpher and Gafni 2001, p 317), or members of the general community (Donaldson 2001). Preferences for a programme or programme attribute may be associated with demographic characteristics. A formal analysis of this variation, either through regression or cluster analysis, may reveal important information about the differing preferences of demographic subgroups (Donaldson 1999, p 557–9). Recall that the minimum sample size for stated preference measures is one (Orme 1998, p 8), and thus we can consider as many subgroups as necessary without resorting to averaging across individuals with different preferences.

Comprehension of the tasks

A significant problem with the assessment of stated preferences is that, under many circumstances, the subjects may be unable to comprehend, or participate in, the various tasks needed for a valid estimation. The researcher needs to ensure that the respondent understands the task as well as possible; otherwise, a spurious valuation may result (for example, valuations based on perceived costs, rather than benefits). It is not clear how well one needs to explain the clinical and scientific foundations of the interventions being valued, given that, in clinical settings, the vast majority of patients will receive less than full information before they are asked to choose among treatments. A rule of thumb may be to provide a similar level of information to the stated preference respondent as one may expect an actual patient to receive, so that the results better resemble revealed preferences (ie no information bias). This approach departs from the classic assumption in economics that decisions are made under perfect information. Most revealed preferences are observed under conditions of less than perfect information, and stated preference should have similar information levels.

Inertia and flexibility of preferences

Another problem associated with measurement of stated preferences is that there are both a degree of inertia and a degree of flexibility in the valuations that respondents may have. It has long been known that individual preferences display an inertia towards the status quo (Porter and MacIntyre 1984, p 11197). This phenomenon can be explained by disutility from change, or the possible ambiguity involved in new interventions (Viscusi and Chesson 1999, p 153). Conversely, preferences are often flexible, being subject to fads and trends, or changes in response to advertising. This may cause problems in the assessment of values in relation to interventions, and could lead to commercial interests advertising their products before they are evaluated. This poses a challenge to both

the economist (who is trained that preferences are static) and the medical scientist (who often wants to estimate some 'true' effect). When it comes to preferences, there is no 'gold-standard', as the preferred option today may not be the preferred option tomorrow. Stated preference methods allow us to measure changes in preferences and consider explanations as to why preferences may change.

Ability to pay

A key limitation of stated preference measures is that they may be affected by ability to pay (Donaldson et al 2002, p 57–9). Those with different incomes have a different perception and/or opportunity cost of money. This limitation only becomes important when one wants to compare WTP across individuals or aggregate across individuals. It is not a new problem in economics (Bridges 1999, p 291).

The ability to pay issue can be addressed in a number of ways. These include weighting the responses, focusing on marginal valuations, or using multiple regression methods (Donaldson 1999, p 554–5; Donaldson et al 2002, p 59). Donaldson et al (2002, p 55–8) argue that income also affects non-monetary valuations (such as those that underpin QALY estimation in CEA), and hence should not be grounds for specifically avoiding stated preference measures.

Nesting of valuations

Given that stated preference tools can be used to value both complete interventions and parts of interventions (either a single or subgroup of the attributes), there may be difficulties in the summing of preferences. For example, the value of a complete intervention may not equal the sum of the values of its components. This issue is known as the 'embedding' or scale problem (ie the inability to perceive and value correctly the differences between small and large effects) (Bateman et al 2002, p 392). These concepts are similar to the traditional concepts of economies of scope and economies of scale, respectively, but are applied to preferences rather than production (Birch and Donaldson 1987). In such circumstances one may need to consider a multiplicative/nonlinear, rather than additively separable, valuation model (Bridges et al 2002, p 47). For example, after spinal cord injury, an individual might have a much higher marginal valuation for the return of motor function in one arm as compared with the marginal valuation of return of motor function to the second. The ideal method for handling these issues is by doing a holistic analysis (ie including all relevant attributes and interventions) and examining preferences including important interaction terms.

Policy implications

Stated preference techniques are more grounded in economic theory than is the more traditional approach of CEA. However, a broader adoption of the methods would raise several important policy questions. These revolve around three key concepts: (1) the ability to pay; (2) the bias towards the status quo; and (3) endogenous preferences.

Should policy makers be concerned with ability to pay?

While there has been much discussion on the effects of income on economic valuation, little attention has been given to resulting policy implications. If an individual values a particular health intervention at lower than its marginal cost, should the intervention be provided for that person? While one may resort to equity weighting, or similar, to inflate the valuations of individuals on limited incomes, this may lead to a less than optimal situation. Consider a middle-aged male requiring coronary artery bypass surgery costing US$40 000, but valuing it at only US$20 000. For completeness we could assume that this is the sum of all individuals' valuations for this particular individual's surgery (ie accounting for all forms of altruism/externalities) and that we have the rare case where WTP = WTA. Is it efficient, just or ethical to provide this service when one could compensate the individual/s for not providing the surgery and have a surplus of US$20 000 that could be used to provide other services? While variances in preferences across income groups can reflect differences in available opportunities, it can also be looked at as differences in opportunity cost. Economists recognise that revealed preferences are affected by income, with associated policy implications. This matter cannot be dealt with by health economists alone. It needs rigorous policy/ethics consideration if we are to make the most of information on stated preferences in policy decisions.

How do policy makers encourage individuals to consider new programmes?

As stated above, there is a tendency for individuals to prefer the status quo. This inertia is important when considering significant changes in policy or provision of health care for three reasons.

1. *The timing of any evaluation is very important.* A novel programme may be valued at less than its 'potential' if it has not been implemented yet, or for a long enough period for individuals to understand it and adjust their consumption and preferences. Here, we can define the *potential value* as the highest possible value after implementation (ie when the individuals' preferences for the intervention have fully developed).
2. *Management of change.* If policy makers want to change the provision of a service or intervention, then they have to manage the change-over period to avoid consumer backlash. A novel intervention can have associated disutility arising from the change itself. This disutility may be minimised through a 'change management' approach, preparing the community for the new intervention, and thus increasing its marginal valuation.
3. *Branding effects*. Policy makers need to account for any possible 'branding' effects (preference for known policies/brands) when interpreting the results of an economic evaluation. Many examples exist of the positive effects of branding on marginal valuations in health care (eg large teaching hospitals branding smaller community hospitals to encourage utilisation).

Can/should policy makers influence decisions?

If policy makers can overcome inertia in preference through a well planned change management strategy, this suggests that people have 'endogenous preferences'. Can policy makers then affect individuals' stated preferences, and hence, valuations, for all types of programmes? If public policy can affect preferences, then static evaluations become redundant. As Bowles (1998, p 75) puts it:

> If preferences are affected by the policies or institutional arrangements we study, we can neither accurately predict nor coherently evaluate the likely consequences of new policies or institutions without taking account of preference endogeneity.

These sentiments are reinforced by Becker (1996, p 6):

> Sub-utility functions of goods do not provide stable foundations because these functions 'shift' over time in response to advertising, addictions, and other behaviors that change personal and social capital.

Becker is advocating the use of a higher order utility function that accounts for the impacts of both time (through personal capital stocks) and of public and corporate policies (through social capital). His 'sub-utility' functions are equivalent to those utility functions that do not allow for endogenous preferences; that is, the type that we have traditionally attempted to measure.

Conclusion

This paper has given an overview of three stated preference methods that are more common in the health economics literature. These, however, are not the only techniques that can be used to estimate a stated preference, and stated preference techniques are only limited by the economic theory that underpins them and the imagination of the investigator. For example, one could simulate the market for health care by giving an individual a budget of $\$M$ and prices of N health care options, with prices at $\$P_n$. By varying M and the various P_n, one would be able to not only estimate demand curves and price elasticities, but cross price and income elasticities.

This paper has also discussed several important limitations and policy implications of stated preference techniques. While there is not a simple solution to the three policy-level issues raised in the above consideration of stated preference techniques, these issues are not avoided by other evaluations such as CEA. Rather, stated preference measures are more transparent and less rigid, thus offering a better insight into the real problems of understanding preferences. No matter which evaluation tool we use, resources will always be scarce, and ability to pay merely implies that some individuals have higher levels of resources than others. If policy makers want to account for ability to pay in their funding decisions, health economists should not force a decision upon them. Rather, health economists should attempt to measure whether economic evaluations in health are affected by ability to pay, so as to better inform policy makers on the potential need for such adjustments.

The issues of inertia and endogenous preferences may offer more scope for contributions by health economists. Input can be provided through the development and implementation of appropriate change management strategies, or through advising policy makers on the appropriate 'social capital' (ie information and/or understanding) needed to make a given policy more valued by the community (Bridges 1999, p 288–9). However, these strategies border on health marketing, rather than health evaluation. Given that stated preference techniques, in one form or another, have always been used by the discipline of

marketing, perhaps the movement towards stated preference tools is symptomatic of the changing role of the health economist away from being a purely objective evaluator.

Acknowledgements

This research was funded, in part, by a grant, 'Current and inter-temporal uncertainty in cost-effectiveness analysis', from AHRQ (HS-01-3738). I would like to thank two anonymous referees and the participants of my class 'Stated preference evaluation' for their comments and suggestions. I would also like to thank Charlene Bridges and Darcey Terris for their comments and corrections.

Notes

[1] An interested reader may benefit from any number of textbooks in the area of stated preferences found in the marketing and environmental management literatures. Currently, there are no comprehensive texts specifically related to the use of state preference techniques in health economics; however, Ryan et al (2001) offers a good overview of a number of various stated preference techniques.

References

Arrow K, Solow R, Portney PR, Leaner EE, Radner R, Schuman H. 1993. Report of the NOAA panel on contingent valuation. *Fed Regist*, 58: 4601–14.

Bateman I, Carson R, Day B, Hanemann M, Hanley N, Hett T, Jones-Lee M, Loomes G, Mourato S, Ozdemiroglu E et al. 2002. Economic valuation with stated preference techniques: a manual. Cheltenham, UK: Edward Elgar. p 1–480.

Becker GS. 1996. Accounting for tastes. Cambridge, MA: Harvard Univ Pr. p 1–268.

Birch S, Donaldson C. 1987. Cost-benefit analysis: dealing with the problems of indivisible projects and fixed budgets. *Health Policy*, 7:61–72.

Birch S, Donaldson C. 2003. Valuing the benefits and costs of health care programmes: where's the 'extra' in extra welfarism? *Soc Sci Med*, 56:1121–33.

Birch S, Gafni A. 1992. Cost effectiveness/utility analysis: do current decision rules lead us to where we want to be? *J Health Econ*, 11: 279–96.

Birch S, Gafni A. 1994. Cost effectiveness ratios: in a league of their own. *Health Policy*, 28:133–41.

Blaug M. 1998. Where are we now in British health economics? *Health Econ*, 7(1 Suppl):S63–78.

Bowles S. 1998. Endogenous preferences: the culture consequences of markets and other economic institutions. *J Econ Lit*, 36:75–111.

Bridges J. 1999. General utility theory. In Baldry J, ed. Economics and health: 1998. Proceedings of the Twentieth Australian Conference of Health Economists. 1999 July 7–8; Sydney. Australian Studies in Health Service Administration Nr 86. Sydney: University of New South Wales. p 281–300.

Bridges J. 2003. Factors in studying patient-physician communication. *JAMA*, 290:2543–4.

Bridges J, Hanson R. 2001. The importance of age and other variables in predicting paediatric patient flows in New South Wales. *Aust Health Rev*, 24:94–9.

Bridges J, Stewart M, King M, van Gool K. 2002. Adapting portfolio theory for investment in health with a multiplicative extension for treatment synergies. *Eur J Health Econ*, 3:47–53.

Coast J. 1999. The appropriate use of qualitative methods in health economics. *Health Econ*, 8:345–53.

Coast J. 2001. Citizens, their agents and health care rationing: an exploratory study using qualitative methods. *Health Econ*, 10:159–74.

Culyer A. 1990. Commodities, characteristics of commodities, characteristics of people, utilities and the quality of life. In Baldwin, S, Godfrey C, Propper C, eds. Quality of life: perspectives and policies. London: Routledge. p 9–27.

Donaldson C. 1998. The (near) equivalence of cost-effectiveness and cost-benefit analysis: fact or fiction? *Pharmacoeconomics*, 13:389–96.

Donaldson C. 1999. Valuing the benefits of publicly-provided healthcare: does 'ability to pay' preclude the use of 'willingness to pay'. *Soc Sci Med*, 49:551–63.

Donaldson C. 2001. Eliciting patients' values by use of 'willingness to pay': letting the theory drive the methods. *Health Expect*, 4:180–8.

Donaldson C, Birch S, Gafni A. 2002. The distribution problem in economic evaluation: income and the valuation of costs and consequences of health care programmes. *Health Econ*, 11:55–70.

Donaldson C, Shackley P. 1997. Does 'process utility' exist? A case study of willingness to pay for laparoscopic cholesystectomy. *Soc Sci Med*, 44:699–707.

Dong H, Kouyate B, Cairns J Sauerborn R. 2003. Willingness-to-pay for community-based insurance in Burkina Faso. *Health Econ*, 12: 849–62.

Elixhauser A, Halpern M, Schmier J, Luce BR. 1998. Healthcare CBA and CEA from 1991–1996: an updated bibliography. *Med Care*, 36 (5 Suppl):MS1–9, MS18–147.

Gafni A. 1991. Willingness to pay as a measure of benefits: relevant questions in the context of public decision making about health care programs. *Med Care*, 29:1246–52.

Gerard K, Mooney G. 1993. QALY league tables: handle with care. *Health Econ*, 2:59–64.

Graff Zivin J, Bridges J. 2002. Addressing risk preferences in cost-effectiveness analyses. *Appl Health Econ Health Policy*, 1:135–9.

Guba EG, Lincoln YS. 1981. Effective evaluation: improving the usefulness of evaluation results through responsive and naturalistic approaches. San Francisco: Jossey-Bass. p 1–423.

Hanemann M. 1999. The economic theory of WTP and WTA. In Bateman I, Willis K, eds. Valuing environmental preferences: theory and practice of the contingent valuation method in the US, EU and developing countries. Oxford: Oxford Univ Pr. p 42–96.

Hoffmann C, Graf von der Schulenberg JM. 2000. The influence of economic evaluation studies on decision making: a European study. *Health Policy*, 52:179–92.

Johannesson M, Jonsson B. 1991. Economic evaluation in health care: is there a role for cost benefit analysis? *Health Policy*, 17:1–23.

Kahneman D, Tversky A. 1982. Prospect theory: an analysis of decisions under risk. *Econometrica*, 47:263–91.

Lawson T. 1995. A realist perspective on contemporary economic theory. *J Econ Issues*, 24:1–32.

Lindholm L, Rosen M, Emmelin M. 1998. How many lives is equity worth? A proposal for equity adjusted years of life saved. *J Epidemiol Community Health*, 52:808–11.

Litva A, Coast J, Donovan JL, Eyles JD, Shepherd M, Tacchi J, Abelson J, Morgan K. 2002. The public is too subjective: public involvement at different levels of health-care decision making. *Soc Sci Med*, 54: 1839–51.

Lomas J. 1998. Social capital and health: implications for public health and epidemiology. *Soc Sci Med*, 47:1181–8.

Louviere JJ. 1988. Analyzing decision making: metric conjoint analysis. Quantitative application in the Social Science Series. Nr 67. Thousand Oaks: Sage Publ. p 1–95.

Louviere JJ, Henley DH, Woodworth G, Meyer RJ, Levin IP, Stoner JW, Curry D, Anderson DA. 1981. Laboratory simulation verses revealed preference methods for estimating travel demand models: an empirical comparison. *Trans Res Rec*, 79:42–51.

Louviere JJ, Hensher DA, Swait JD. 2000. Stated choice methods: analysis and application. Cambridge: Cambridge Univ Pr. p 1–418.

McFadden D. 1974. Conditional logit analysis of qualitative choice behaviour. In Zarembka P, ed. Frontiers in econometrics. New York: Acad Pr. p 105–42.

McIntosh E, Donaldson C, Ryan M. 1999. Recent advances in the methods of cost benefit analysis in healthcare: matching the art to the science. *Pharmacoeconomics*, 15:357–67.

Mooney G. 1994. What else do we want from our health service? *Soc Sci Med*, 39:151–4.

Mooney G. 1998. Communitarian claims as an ethical basis for allocating health care resources. *Soc Sci Med*, 47:1171–80.

Mooney G, Lange M. 1993. Anti-natal screening: what contributes to benefit? *Soc Sci Med*, 37:873–8.

Morrison GC, Gyldmark M. 1992. Appraising the use of contingent valuation. *Health Econ*, 12:32–43.

O'Brien B, Gafni A. 1996. When do the dollars make sense? Towards a conceptual framework for contingent valuation studies in health care. *Med Decis Making*, 16:288–99.

Olsen JA, Donaldson C. 1998. Helicopters, hearts and hips: using willingness to pay to set priorities for public sector healthcare programmes. *Soc Sci Med*, 46:1–12.

Orme B. 1996. Which conjoint method should I use? Sawtooth Software Research Paper Series. Sequim, WA: Sawtooth Software.

Orme B. 1998. Sample size issues for conjoint analysis. Sawtooth Software Research Paper Series. Sequim, WA: Sawtooth Software.

Patton MQ. 1987. How to use qualitative methods in evaluation. Thousand Oaks: Sage Publ. p 1–176.

Patton MQ. 1990. Qualitative evaluation and research methods. 2nd ed. Thousand Oaks: Sage Publ. p 1–532.

Porter M, MacIntyre S. 1984. What is, must be best: a research note on conservative or deferential responses to antenatal care provision. *Soc Sci Med*, 19:11197–200.

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

Richards TJ, Richards L. 1994. Using computers in qualitative research. In Denzin NK, Lincoln YS, eds. Handbook of qualitative research. Thousand Oaks: Sage Publ. p 445–62.

Robinson R. 1993. Cost benefit analysis. *BMJ*, 307:924–6.

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

Ryan M, Gerard K. 2003. Using discrete choice experiments to value health care programmes: current practice and future research reflections. *Appl Health Econ Health Policy*, 2:55–64.

Ryan M, Scott DA, Reeves C, Bate A, van Teijlingen ER, Russell EM, Napper M, Robb CM. 2001. Eliciting public preferences for healthcare: a systematic review of techniques. *Health Technol Assess*, 5:1–186.

Sculpher M, Gafni A. 2001. Recognising diversity in public preferences: the use of preference sub-groups in cost-effectiveness analysis. *Health Econ*, 10:317–24.

Seidel J. 1991. Method and madness in the application of computer technology to qualitative data analysis. In Fielding NG, Lee RM, eds. Using computers in qualitative research. London: Sage Publ. p 107–16.

Skjoldborg U, Gyrd-Hansen D. 2003. Using conjoint analysis: the cost variable: an Achilles' heel? *Health Econ*, 12:479–91.

Sorrell JM, Redmond GM. 1995. Interviews in qualitative nursing research: differing approaches for ethnographic and phenomenological studies. *J Adv Nurs*, 21:1117–22.

St John W, Johnson P. 2000. The pros and cons of data analysis software for qualitative research. *J Nurs Scholarsh*, 32:393–7.

Stalhammar NO. 1996. An empirical note on willingness to pay and starting-point bias. *Med Decis Making*, 16:242–7.

Stinnett AA, Mullahy J. 1998. Net health benefits: a new framework for the analysis of uncertainty in cost-effectiveness analysis. *Med Decis Making*, 18:S68–80.

van der Pol M, Ryan M, Donaldson C. 2003. Valuing food safety improvements using willingness to pay. *Appl Health Econ Health Policy*, 2:99–107.

Viscusi WK, Chesson H. 1999. Hopes and fears: the conflicting effects of risk ambiguity. *Theory Decis*, 47:153–78.