Ordinal preference elicitation methods in health economics and health services research: using discrete choice experiments and ranking methods

Shehzad Ali* and Sarah Ronaldson
Department of Health Sciences, University of York, Heslington, York YO10 5DD, UK

Introduction: The predominant method of economic evaluation is cost–utility analysis, which uses cardinal preference elicitation methods, including the standard gamble and time trade-off. However, such approach is not suitable for understanding trade-offs between process attributes, non-health outcomes and health outcomes to evaluate current practices, develop new programmes and predict demand for services and products. Ordinal preference elicitation methods including discrete choice experiments and ranking methods are therefore commonly used in health economics and health service research.

Areas of agreement: Cardinal methods have been criticized on the grounds of cognitive complexity, difficulty of administration, contamination by risk and preference attitudes, and potential violation of underlying assumptions. Ordinal methods have gained popularity because of reduced cognitive burden, lower degree of abstract reasoning, reduced measurement error, ease of administration and ability to use both health and non-health outcomes.

Areas of controversy: The underlying assumptions of ordinal methods may be violated when respondents use cognitive shortcuts, or cannot comprehend the ordinal task or interpret attributes and levels, or use ‘irrational’ choice behaviour or refuse to trade-off certain attributes.

Current use and growing areas: Ordinal methods are commonly used to evaluate preference for attributes of health services, products, practices, interventions, policies and, more recently, to estimate utility weights.

Areas for on-going research: There is growing research on developing optimal designs, evaluating the rationalization process, using qualitative tools for developing ordinal methods, evaluating consistency with utility theory, appropriate statistical methods for analysis, generalizability of results and comparing ordinal methods against each other and with cardinal measures.
Introduction

Health economic evaluation involves valuation of health states, services, products, interventions, practices and policies. The predominant methods of economic evaluation in the literature are cost-effectiveness analysis (CEA) and cost–utility analysis (CUA). In the case of CEA, the outcome measure is programme-specific, such as the reduction in blood pressure, number of positive cases detected, change in asthma episode days and gain in life years. CEA can be useful when the decision maker is interested in comparing alternatives within a particular field, for instance, cancer screening. However, for a decision maker with broad health sector mandate, the more appropriate method of economic evaluation is CUA because it uses a generic measure of health outcome that can be compared across programmes. The most commonly used outcome measure in CUA is the quality-adjusted life year (QALY) that combines survival and health-related quality of life (HRQoL). The QALY measure is preference-based, i.e. it incorporates preference or desirability of the general public for health-related consequences (such as health states) of the alternative programmes being evaluated. The commonly used preference elicitation methods used for QALY-based analysis are the standard gamble (SG) and time trade-off (TTO) (see Whitehead and Ali for further details on the QALY-based approach to economic evaluation).

While the QALY-based approach is attractive because of its generic nature, it is not suitable for all evaluation studies. For instance, when comparing health-related programmes (in this context, the term ‘programme’ is used in a generic sense to represent health-related services, products, interventions, practices and policies), the traditional QALY approach assumes that consumer [i.e. individuals whose welfare is being evaluated, including patients, service providers, carers, decision makers, general public and other consumers of health-related programmes] preference is only influenced by outcomes that have a direct impact on HRQoL and/or survival. This can be a limitation since process characteristics (or attributes) and non-health outcomes can play a crucial role in consumer preference and may be of interest for the decision maker evaluating or developing health-related programmes. Such process attributes may include waiting time to access healthcare, choice of medical staff, choice of service level and type.
of provider, distance to health facility, time of appointment, availability of home visits, cleanliness of wards and toilets, mode of treatment administration, frequency of medication and out-of-pocket cost. Similarly, the non-health outcomes that may influence consumer preference may include the impact on education, dignity, safety and ability of social participation, availability of non-health-related services, control over daily life and impact on family members. The traditional QALY-based approach ignores such attributes and does not lend itself directly to establishing trade-offs between health outcomes, non-health outcomes and process attributes. Moreover, in many situations, the QALY is not the direct outcome of interest for comparing health programmes. For instance, a decision maker may be interested in evaluating the relative value of characteristics (or attributes) of health-related programmes (such as attributes of health insurance packages) to predict the demand for services with different specifications. Here, the QALY is not the direct outcome of interest; rather it is the relative preference for process and outcome attributes that are of interest (such as waiting time to access health services, type and level of services covered, reimbursement of service fees and level of insurance contribution). Other examples of situations where QALY-based approach is not appropriate include evaluation of relative preferences for attributes of treatments (such as mode of drug administration, frequency of use, clinical efficacy, side effects and cost), or establishing willingness-to-pay for healthcare products or services (such as out-of-hours clinics). In such situations, alternative preference elicitation methods are required to directly compare attributes of alternative programmes.

The two commonly used alternative preference-based approaches in health economics are discrete choice experiments (DCEs) and ranking methods. These approaches are commonly used to evaluate how individuals value and trade-off characteristics of health-related programmes, to predict the demand for services or products with various characteristics and to establish willingness-to-pay for programmes with different characteristics. Such information may be crucial to decision makers for evaluating current practices and developing new products and services based on relative preference for levels of health, non-health and process attributes. Examples include evaluating patient preference for treatment of diabetes, community preference for health service providers, consultants’ or service providers’ preference for job characteristics, public preference for resource allocation decisions in the healthcare system, healthcare decision makers’ attitude towards trade-off between equity and cost-effectiveness, community preference for insurance packages, preference of general population for long-term care and patients’ willingness-to-pay for alternative...
services.\textsuperscript{16,18,19} DCE and ranking studies can be used as stand-alone studies, or as part of full economic evaluations.\textsuperscript{1,29}

The aim of this article is to introduce the concepts and methods of DCEs and ranking studies, and to critically and comprehensively appraise these methods in relation to their use in economic evaluation and health services research. This article is structured as follows: first, we will introduce the concept of preference elicitation methods, followed by a discussion on the need for alternative elicitation methods. Then, we will introduce and critically appraise DCE and ranking studies and discuss the future research frontier for these preference elicitation methods. Further references are provided to direct the interested reader.

**What are preference elicitation methods?**

Economists assume that individuals prefer one health programme (or health state) over another if they derive more satisfaction (utility) or less dissatisfaction (disutility) from it. For instance, when presented with choice between two treatments that differ in terms of levels of process and outcome attributes (such as level of efficacy, mode of administration and level of risk of adverse events), individuals choose the alternative that provides them with greatest utility. While utility of each choice alternative is not observed by researchers, the choice decisions made by individuals can reveal the underlying utility or value they associate with each alternative. This forms the basis for the use of utility-based preference methods. These methods include ‘stated preference elicitation methods’ and ‘revealed preference methods’. The former ask individuals to state their preference from a set of two or more alternatives in a hypothetical situation. Stated preference elicitation involves trade-offs being made which reveal the value (or utility) that individuals associate with health programmes or attributes of health services and products.

On the other hand, revealed preference methods involve the valuation of alternative programmes by evaluating actual consumer behaviour in real-life market settings, which can in turn reveal the trade-offs that individuals actually made.\textsuperscript{36} For instance, an analysis of actual market data to evaluate individual choices of dental health plans in real life can reveal how individuals trade-off different attributes of health plans to maximize their utility. However, conducting valuations based on revealed preferences is restricted by the actual existing markets for goods and services (or in our case, health-related programmes); this poses limitations, particularly in the health sector where open markets for goods and services (such as antibiotics for infections) often do not
exist. Also, in most countries with public and private insurance, consumers rarely face market prices for goods and services. Moreover, revealed preference methods suffer from problems related to lack of consumer knowledge about products and services (including incomplete disclosure about healthcare products/services) or subsidised price at the point of use. Also, it is difficult to imagine how revealed preference can be used for valuation of health states that are not usually chosen voluntarily. (Individuals usually do not choose their own health state, i.e. there is no market for health states, although individuals may choose the process that leads to a health state, for instance, poor diet.) Finally, revealed preference methods are not useful for evaluating hypothetical scenarios or interventions that are under development and policies for which market data do not exist. Hence, ‘stated preference’ methods are widely used in the healthcare field.

Stated preference methods are especially useful when a product, service or policy is not currently available in real life, or where the aim is to evaluate preferences across currently available alternatives. Stated preference methods can help to understand the mechanism of the choice process and the trade-offs that individuals are willing to make; this is done by creating hypothetical choice scenarios with different levels of attributes. (Note that some studies have criticized stated preference methods on the grounds that they do not represent the actual choice decisions. Some critics of stated preference methods argue that individuals may behave differently when faced with actual choice decisions in real life markets compared to hypothetical scenarios in stated preference exercises.) It should be noted that some studies have combined the revealed and stated preference approaches. However, most of these studies have been conducted outside of the health economics field.

Stated preference elicitation methods can be understood as either cardinal or ordinal (Fig. 1). Cardinal methods generate preferences in a quantitative form based on answers from respondents, i.e. they give direct estimates of the degree to which one health programme (or health state) is preferred over another. The most common cardinal methods include the SG and TTO. During these tasks, individuals trade-off between probabilities, uncertainties or risks associated with health states or services. The methods are most often used to value health states by asking individuals to trade-off years of perfect health (TTO) or risk of immediate death (SG) against remaining in a particular health state for a period of time. Cardinal measures also include the visual analogue scale (VAS) which is a rating exercise to value health status on a scale between 0 and 100. The VAS has been criticized on theoretical and empirical grounds and is less frequently used as a valuation tool. In health economics, cardinal methods are primarily
used to produce quantitative weights (i.e. utility values) for health states that are subsequently used to calculate QALYs. See Whitehead and Ali\textsuperscript{6} for a detailed account of cardinal methods in economic evaluation.

Ordinal preference elicitation methods, in contrast to cardinal methods, are concerned with ordering preferences of two or more alternatives without directly establishing the degree of preference of one alternative over the other. Commonly used ordinal preference methods include DCEs and ranking exercises.\textsuperscript{45} These methods are relatively easier for respondents to understand and also easier to administer compared with SG and TTO. The methods have sometimes been used to value health states, but more commonly to evaluate healthcare services, products, practices, interventions and policies.\textsuperscript{46} Ordinal methods can also be used to evaluate willingness-to-pay or willingness-to-accept for a change in the level of an attribute of a service or product.\textsuperscript{47} Ordinal measures can be transformed into cardinal measures for use in economic evaluation\textsuperscript{48} or are useful as stand-alone valuation tools in their own right.

**Why use alternative preference elicitation methods?**

The most appropriate valuation method (cardinal or ordinal) is debated; methods differ in relation to their mode of administration, ease of comprehension, underlying assumptions, scale of measurement, nature of trade-offs, use of props and diagrams, time allowed for reflection and individual versus group interviews.\textsuperscript{45} Different elicitation...
methods can generate different health valuations or preferences, with there being substantial evidence to suggest that this is the case.\textsuperscript{43} However, there has been no consensus towards the most preferred method.\textsuperscript{49} Some studies argue that the more commonly used valuation methods, such as the SG or TTO, may not be the most appropriate or feasible in all situations. Brazier and Ratcliffe\textsuperscript{45} pointed out that the SG, which has often been described as the gold standard for outcome measurement, may in practice violate the underlying theoretical axioms of expected utility theory (EUT) (EUT postulates that individuals choose between alternatives to maximize their expected utility. Expected utility of an alternative is calculated by estimating the utility of each possible outcome and then multiplying their probability of occurrence, and finally summing up to calculate the expected utility of the alternative. These expected utilities are then compared to maximize the utility gain), which SG is rooted in.\textsuperscript{50,51} In addition, the values generated by the SG can be contaminated by risk attitude (i.e. depending on how risk averse individuals are) and loss aversion (i.e. tendency to strongly prefer avoiding losses to acquiring gains); similarly, the TTO can be contaminated by time preference (i.e. individuals have different (usually higher) preference for health now over future health), duration effects and possibly unwillingness of some respondents to sacrifice any of their life expectancy.\textsuperscript{52} Bleichrodt\textsuperscript{53} has explored why people may violate expected utility and the direction of bias associated with SG and TTO methods.

Ordinal methods overcome some of the biases of SG and TTO methods and produce more consistent responses than TTO and SG.\textsuperscript{42} These methods have therefore become increasingly popular in recent years.\textsuperscript{54,55} They offer advantages in the form of relative ease of comprehension and administration, in addition to greater reliability in terms of reduced measurement error.\textsuperscript{45} There are practical advantages for the use of ordinal methods over more commonly used methods such as SG and TTO, including reduced cognitive burden posed to respondents and a lower degree of abstract reasoning, which are particularly important in settings with limited educational and numeracy attainment.\textsuperscript{52} For this reason of lower complexity, ordinal methods are expected to more accurately reflect decision maker preferences.\textsuperscript{56} For example, Coast\textit{et al.}\textsuperscript{57} point out the questionable appropriateness of complex tasks such as the SG and TTO for an older population group. It is less complex for respondents to provide answers about their preference for health state X versus health state Y (ordinal techniques), rather than needing to quantify the magnitude of the preference for one scenario over the other (cardinal techniques). Moreover, some cardinal tasks may cause distress to the respondents, particularly when there is reference to early death. Finally, cardinal methods are primarily
useful for evaluation of health states; they have found limited use in understanding trade-offs between health and non-health outcomes (or processes) that is crucial to evaluation of health services and products. A great advantage of ordinal methods is that they allow investigation of such trade-offs. In summary, the concerns over the use of cardinal methods, the difficulty of administration of SG and TTO methods and the desirable properties of ordinal methods have contributed towards the increasing recent use of ordinal techniques.57

What are ordinal preference elicitation methods?

The main ordinal stated preference methods are summarized below (Fig. 1).

**Discrete choice experiments**

DCEs ask individuals to choose between two or more alternatives, where each alternative is usually described using two or more attributes (see the Discrete choice experiment section). The alternatives to choose from may include health states, health services, products, practices, interventions or policies. The attributes can be process or outcome attributes and may or may not be health-related (for instance, cure rates and waiting times).

**Ranking**

A ranking exercise involves individuals being asked to place in order (partial order or complete order) a set of attributes or alternatives according to which they consider to be the best to the worst, for example. As in the case of DCEs, the attributes can be health or non-health-related.

**Ordered categorical responses**

An alternative method of valuation is the use of an ordered categorical response scale. This approach involves defining a set of response alternatives, followed by asking respondents to rate each health state individually according to how good or bad a respondent considers the value associated with that alternative to be. Hence, no direct comparison is made between alternatives, which is different from ranking. This approach is relatively less commonly used in health economics;
hence, the focus of further discussion will be on the first two approaches.

A useful document providing good practice guidelines for conducting ordinal preference studies is a report produced by the International Society for Pharmacoeconomics and Outcomes Research (ISPOR) Good Research Practices for Conjoint Analysis Task Force. The document provides an impetus for good quality applied research; it also includes a useful checklist for researchers interested in conducting or appraising ordinal preference methods.

**History of ordinal preference elicitation methods**

Until recent years, the main purpose of ordinal ranking techniques was to provide a ‘warm-up’ exercise, in order to prepare respondents for the main task of cardinal valuation. However, the use of ordinal preference methods, either as stand-alone exercises or as part of cost-effectiveness or cost–benefit analysis, has become common in recent years. This is partly attributable to the strong methodological foundation that exists for estimating cardinal values from ordinal information, primarily originating in psychology but also commonly applied in areas of environmental economics, transport research, consumer marketing and political science. This has facilitated much wider use of ordinal methods in economic evaluation.

Work on ordinal preference elicitation methods can be traced back to 1920s, with work in the area being pioneered by Thurstone who proposed the law of comparative judgement and the random utility theory (RUT) that laid the foundations of the process of pairwise comparisons. Thurstone’s paired comparison approach was applied to estimate health valuations by Fanshel and Bush. Other approaches to modelling ordinal data, such as the Bradley–Terry model, have also been proposed. This approach proposed the use of logistic functional form instead of the normal density function, which facilitated valuation functions to be estimated from ordinal data. Kind subsequently compared the Thurstone and Bradley–Terry models for scaling health indicators.

Louviere and Woodworth pioneered the multi-attribute DCEs by proposing the use of experimental design theory to construct choice sets of profiles (or alternatives). This work was based on the multi-attribute utility theory and was found to be consistent with the conditional logit models proposed by McFadden. More recently, ordinal methods, such as DCEs, which have roots in RUT, were introduced to health economics by Propper and advocated by Ryan and colleagues. It should be noted that in the initial years, the early role of
DCEs was to go beyond the conventional QALY-based approach to incorporate process characteristics and non-health outcomes to understand the trade-offs between levels of attributes of health services or goods. However, in more recent years, it has been recognized that ordinal preference methods can contribute to both the CEA and cost–benefit analysis. As a result, valuation of health states and process outcomes has been investigated in several studies. When used for valuation of health states, ordinal methods may compare health states defined by generic standardized multi-attribute descriptive systems such as the EQ-5D, SF-6D or the HUI, or condition-specific vignettes. See Louviere and Lancsar for a detailed history of choice experiments in health.

Discrete choice experiment

The DCE is the most common type of ordinal preference method used in health economics and health services research. Carson and Louviere defined the DCE as ‘a general preference elicitation approach that asks agents to make choice(s) between two or more discrete alternatives where at least one attribute of the alternative is systematically varied across respondents in such a way that information related to preference parameters of an indirect utility function can be inferred’. Less formally, a DCE can be understood as a survey-based method to evaluate the relative importance of health states or attributes of services, interventions, practices and policies to establish the trade-offs that individuals are willing to make between levels of attributes in order to maximize their utility from the available alternatives. This approach can then estimate the value of given alternatives to the respondents.

In the literature, several generic terms have been used to refer to DCEs. These include, but are not limited to, conjoint analysis, conjoint experiments, choice experiments, paired comparisons and stated preference methods. However, these terms do not always refer to the same approach and there is much debate about their correct use. See Carson and Louviere for a detailed discussion on nomenclature of stated preference methods.

The use of DCE in health economics and health services research

The DCE approach has been used in several areas of health economics and health services research. We have provided several examples in the Introduction section of the kind of questions that can be answered
using DCEs. Broadly speaking, DCEs can be used to elicit preferences for health-related products and services, to value health outcomes and health states, to quantify trade-offs between attributes, to evaluate non-health outcomes (for instance, GP remuneration package) and to predict expected uptake of policies and products. In effect, any research question that requires preference elicitation based on hypothetical choice scenarios can be addressed using DCEs. Bekker-Grob et al. found that between 1990 and 2008, there was limited focus on valuing health states using DCE methods. However, more recently, DCEs have also been used to estimate cardinal values on the health utility scale for generic and condition-specific health states.

As mentioned earlier, DCEs are also useful tools to understand the trade-offs between health and non-health outcomes that may be difficult to study using the traditional cardinal methods. A typical example is the use of cost as one of the attributes in the experiment to investigate WTP of service users for attributes of services and interventions. Finally, interest has been growing in the use of the DCE approach to explore the equity dimension of health and healthcare. One such study was Lancsar et al. that used the DCE method to derive distributional weights for QALY gains using society’s health values. Such an approach can be used to understand equity-efficiency trade-offs.

**Generic and alternative-specific DCEs**

For practical purposes, DCEs can be understood as either ‘generic’ or ‘alternative-specific’. A generic DCE involves comparing alternatives that are unlabelled (for instance, comparing healthcare providers A and B that are different only in terms of the specified attributes and levels), while an alternative-specific DCE compares labelled alternatives that have inherent meaning to the respondents (for instance, chiropractor or physiotherapist, both of which are then defined further in terms of, say, waiting time, distance and costs). The advantage of using alternative-specific labels is respondent familiarity with the context which reduces the cognitive burden; however, the disadvantage is that respondents may partly ignore the trade-offs between the specified attributes in the DCE and instead focus more on the unspecified attributes of the alternative. The generic design is far more common in the health economics and health services literature. Hence, the focus of the remaining part of this section is on generic DCEs. However, the underlying principles of the two designs are the same.
A hypothetical example of DCE

In a typical DCE study, the context of the experiment is described to the respondents at the start of the survey questionnaire. For instance, in a study evaluating patients’ preferences for the out-of-hours general practice, the respondents will be described the context and setting of the service. Based on this context, individuals are presented with one or more choice questions (or choice sets), each with two or more hypothetical alternatives (or scenarios) to choose from. Each alternative can be completely described by a set of attributes. For example, a DCE survey evaluating treatment strategies in primary care for a particular condition may have 10 choice questions, each with alternatives A and B. Each alternative is then defined by three attributes, i.e. cure rates, waiting time and cost to the patient. The alternatives will differ in terms of the ‘level’ of one or more attributes. For example, in choice question 1, the hypothetical alternative A may offer a cure rate of 70%, waiting time of 1 week and personal cost of £50. Against this, treatment alternative B may offer a better cure rate of 90%, longer waiting time of 2 weeks and higher cost of £75. Respondents will then be asked to choose between alternatives A and B. This would require them to make a trade-off between a better cure rate in treatment B against shorter waiting time and lower personal cost of treatment A. The process is then repeated for the 10 choice questions in the survey with varying levels of the three attributes. In the process of making these choice decisions, individuals reveal the relative significance (preference) of levels of attributes in their decision-making process to maximize their utility.

The number of attributes and choice scenarios (questions) vary between studies. A recent survey of DCEs found that most DCE studies use 2–6 attributes (but can have more) that typically include some of the following: aspects of healthcare products, services and practices, monetary measures, time, level of associated risk or health status domains. However, any other attributes that are relevant to the research question can be included in the study. The number of choice scenarios depend on the study design (discussed later). The above-mentioned survey found that most studies used ≤16 scenarios per respondent. An example of a DCE study from Scott et al. is presented in Figure 2.

Designing the DCE

A DCE study involves several stages, including conceptual development, design, conduct and statistical analysis. The first stage involves defining the research aims of the study and conceptualizing...
the choice process. For example, in a DCE study of out-of-hours care, Scott et al.\textsuperscript{11} specified that the aims of the study were to elicit preferences of the community for different models of out-of-hours care, and to evaluate the relative importance of attributes of out-of-hours care. Once the research objectives are established, the choice process is conceptualized which involves considering the decision-making context for individuals making the choice, evaluating the alternatives currently available (and those that may become available or are being developed) and investigating the attributes that are likely to be important to users or decision makers and/or those that may be amenable to policy change. This process is likely to involve some or all of the following: literature review, individual and focus group interviews with patients, carers, patient groups, local organizations and the general population,
expert clinical opinions, and discussions with policy-makers, senior officials and service providers. As relevant, the investigators should ensure appropriate representation of opinions, including those from diverse socio-demographic groups. Relevance of alternatives to respondents, and the empirical context of products and services should be considered. It is also important to consider whether individuals are familiar with the decision-making context or may require specialized knowledge to be able to make an informed decision. Through this process, important attributes relevant to the decision-making context are selected. Following this, levels of the attribute are selected based on the range of values or situations that each attribute may take within the context of the research question. This is also informed by the above-mentioned process of literature reviews and qualitative research. It is important to be aware that the number of attributes and levels should be balanced against what is cognitively feasible for respondents to handle in a DCE.

Based on attributes and levels, hypothetical alternatives are produced and grouped into choice questions to form the survey questionnaire. A full factorial survey design would include all possible combinations of attributes and levels to generate all possible alternatives ($x^y$, where $x$ is the number of levels, $y$ the number of attributes). However, this usually results in a large number of choice questions that would be too tedious for the respondents and also too expensive to administer. Hence, experimental design methods are used to create smaller and more manageable sets of alternatives to generate fractional factorial designs. A recent review of DCEs in health economics found that all DCEs conducted between years 2001 and 2008 used a fractional factorial design. For an efficient design, it is important to aim for orthogonal and balanced design, i.e. the attributes should be statistically independent of each other, and they should occur equally frequently in the survey. It is common practice to use software (including SPEED, SPSS, SAS and Sawtooth) to develop most efficient designs.

Analyzing the DCE

DCE can be understood in the light of the two underlying theories, i.e. Lancaster’s theory of value, and the RUT. Lancaster’s theory of value argues that individuals view a commodity or service as a combination of several component attributes (i.e. its characteristics) such that the total utility (satisfaction or desirability) of a commodity is a function of separate partial utilities associated with composing attributes of the commodity. The RUT argues that the utility value that an individual attaches to an alternative in a choice scenario cannot be observed by a...
researcher, but it can be summarized by two components: an explainable (or systematic) component and an unexplainable (or random) component.\textsuperscript{64} The explainable component comprises the attributes of alternatives and the individual characteristics, while the unexplainable component comprises the unobserved factors that influence the choice. It is assumed that respondents know the nature of their utility function. However, researchers do not observe this function, but they observe the choices made by individuals. Based on the principle of utility maximization, we assume that an individual will choose alternative A over B in a choice question if the utility derived from A is more than that from B. This ordinal choice decision is used to derive a cardinal valuation by making the assumption that choices over sets of attributes are related to underlying latent cardinal values that are distributed around mean levels for each item. Then a choice-based statistical model is employed to analyse the influence of attribute levels on choice decisions. Subsequently, the relative significance of attributes can be evaluated by taking a ratio of the coefficients of each possible pair of attributes. This provides the marginal rate of substitution (MRS), i.e. the rate at which an individual is likely to give up one level of an attribute in exchange for another attribute.

DCEs are commonly analysed using logit or probit statistical models. A literature review of DCE studies published between 1990 and 2008 found that most studies used McFadden’s multinomial (conditional) logit or random effects probit models.\textsuperscript{46} However, more recently, mixed logit models have gained popularity in this area. These models take account of preference heterogeneity.\textsuperscript{83} It is also common practice to evaluate the interaction between respondent-level characteristics and attributes in the DCE in predicting choice decisions. This relationship is explored by introducing interaction terms in the statistical model.

**Challenges and limitations of DCEs**

Potential methodological limitations of DCEs have been identified. Ryan \textit{et al.}\textsuperscript{84} conducted a qualitative think aloud study to understand the rationalizing process used by respondents while making their choices. They found that many respondents concentrated primarily on one or two attributes to make their choices as this simplified the task for them. Lloyd\textsuperscript{43} adds that individuals employ cognitive heuristics (or shortcuts) to reduce the cognitive burden of the task. As a result, they may be less willing to accept trade-offs for certain attributes that were identified by them to simplify the task. This is formally known as non-compensatory decision-making.\textsuperscript{71} This in turn would have an impact...
on the validity of the results of DCE. Hence, understanding the decision-making process is crucial in the development of a DCE.

Another challenge is that respondents may make judgements about the quality of an intervention (above and beyond what is specified in the choice question) based on known attributes or past experiences. For instance, if a respondent believes that costlier alternatives tend to have better quality, then they may prefer the more costly options even when all options are qualitatively equivalent in terms of the specified attributes. Another challenge is that respondents may have difficulty understanding risk information in DCEs (for instance, the risk of an adverse event). Also, individuals may see events as more likely if they are familiar; respondents also tend to value hazards more highly for other people than themselves.

Some authors have challenged that the conventional methods used to derive welfare measures from DCEs (in particular, willingness-to-pay estimates) may not be consistent with the underlying RUT; these could potentially be improved by using alternative methods of deriving welfare measures. Similarly, Flynn et al. argue that when valuing health states using DCE, some of the underlying assumptions may be violated if respondents refuse any health state to be worse than death. Also, stability of individual preference over time has been debated in the literature, which can potentially have an impact on the inferences drawn from DCEs.

Another challenge is that respondents may not interpret attributes and levels in the way the researchers intended which will have an impact on the interpretation of the valuation estimates from DCEs. Quite often, DCEs include one or more ‘consistency checks’ to assess rationality and internal validity of responses. One of these checks is the dominance test which refers to choice questions in which one of the alternatives is clearly dominant over the other (i.e. at least one attribute in the dominant alternative is better than the dominated and no attribute is worse). Individuals who fail this test are often referred to as ‘irrational’. Another check is the test–retest reliability check which is conducted by repeating a choice scenario later in the survey to check for consistency of responses. While some investigators prefer to drop the respondents from analysis who failed one or more checks, others argue that removal of ‘irrational’ responses may also result in removal of valid responses. It has also been argued that ‘irrational’ behaviour may imply that choice behaviour in DCEs may not be consistent with the utility theory. This is an ongoing area of research. Some studies have also criticized DCEs for being too restrictive in terms of inclusion of attributes. However, albeit its limitations and challenges, the DCE is a fast growing area for applied economists and there will inevitably be further advances in the field in the years to come.
Ranking exercises

It has been argued that some of the limitations of DCEs can potentially be overcome by using ranking methods. For instance, Flynn et al.\(^5^4\) suggest that comparative inference about attributes such as ‘waiting time for an appointment’ is more important to patients than continuity of care cannot be drawn directly from DCEs because DCEs produce comparative estimates for levels of attributes (i.e. scale values) and not for attributes per se. Moreover, some authors argue that the ‘pick one’ nature of the tasks in DCE may be less efficient than some of the ranking exercises. Ranking data can provide more statistical information than choice experiments.\(^5^4\) However, like DCEs, there are limitations of ranking exercises, some of which will be discussed below.

Currently, ranking exercises are less commonly used in health economics than DCEs, partly because these are relatively new to the discipline. It should be noted that some authors regard ranking methods as a type of DCE exercise.\(^6^9\) However, for empirical purposes, we prefer to review them separately to distinguish these methods. Two of the more common ranking approaches include complete ranking exercise and the best-worst scaling (BWS) method.

Complete ranking exercise

The complete ranking exercise is sometimes referred to as the contingent ranking method, especially when cost is one of the attributes varied across alternatives. [Note that the contingent ranking approach is different from the contingent valuation method (not discussed in this article); the latter directly ask individuals about their willingness-to-pay or willingness-to-accept compensation for public goods and services.] The exercise requires respondents to rank order all available alternatives from the most preferred to the least preferred option, providing complete ordering of all alternatives. Fok et al.\(^8^8\) argue that complete ranking may produce more efficient preference estimation compared with DCEs. However, they point out that the sorting task may be cognitively too complex or time-consuming for respondents, resulting in biased preference estimates. The complexity of the task increases as the number of alternatives increases.\(^6^9\) One potential solution is to take into account the ranking capabilities of individuals by including individual-specific functions in the statistical model that describe ranking capabilities of respondents. Fok et al.\(^8^8\) claim that such models would perform well when at least some respondents are able to rank more than one alternative. However, the validity of such models needs to be further investigated. Another potential solution is to use only a
small number of alternatives that are cognitively possible for the respondents to manage. For instance, Slothuus et al.\textsuperscript{89} constructed four cards each containing five attributes with different levels to establish the willingness-to-pay for alleviation of rheumatoid arthritis symptoms. Similarly, Ratcliffe et al.\textsuperscript{42} elicited preferences for 11 sexual health states using a ranking exercise. However, the use of limited alternatives may not always be appropriate and may adversely impact on statistical estimation, especially when several attributes and levels are being evaluated.

Another potential problem with complete ranking is that respondents may find it harder to distinguish between less-preferred alternatives and in turn produce biased ranking at the bottom end. To overcome this, Chapman and Staelin\textsuperscript{90} have suggested using only the first few ranks for estimation. Others have suggested two-stage approaches whereby the respondents may exclude the alternatives they do not prefer and then rank order the remaining alternatives for estimation purposes.\textsuperscript{91} However, there is no consensus on whether these solutions can overcome the limitations of complete ranking exercises. Finally, McCabe et al.\textsuperscript{68} note that some of the underlying assumptions of complete ranking exercises may be unrealistic, resulting in incorrect inferences.

**Best-worst scaling**

Another approach to ranking that is far more popular than complete ranking in health economics is the BWS method. This method was developed by Louviere and Woodworth,\textsuperscript{92} and was empirically used by Finn and Louviere\textsuperscript{93} and further developed by Marley et al.\textsuperscript{94} There are three types of BWS methods, including the ‘object case’, the ‘profile case’ and the ‘multi-attribute case’ (see Flynn\textsuperscript{95} for a detailed discussion with examples). The object case asks respondents to choose best and worst attributes (without levels). The profile case asks to choose the best and worst of the attributes with levels, while the multi-attribute case (closest to traditional DCEs) asks to choose best and worst profiles (alternatives) from a set of multi-attribute profiles. The most common type used in health economics is the profile case BWS. In this method, respondents are presented with one scenario (or profile) at a time. Each scenario has three or more attributes with different levels, and respondents are asked to select the best and worst attribute levels within each scenario, rather than between scenarios as is the case in DCEs. In effect, respondents are asked to pick a pair of attribute levels that are furthest apart on the latent utility scale.\textsuperscript{54} For instance, Potoglou et al.\textsuperscript{22} investigated individual’s social care-related quality of life on nine domains (or attributes) such as food and drink,
accommodation, social participation, control over daily life etc. Each attribute had 2–4 levels. In the choice task, the respondents were presented with several scenarios, one at a time, where each scenario had all nine attributes with one level. The respondents then trade-off between attributes and levels to state best and worst attribute levels, in turn revealing their underlying utility function. BWS has been used for valuation of health states and health interventions. This approach can give additional information to that available from a DCE as it uses the utility of a single level of an attribute as a benchmark rather than an entire scenario.

Empirical studies using BWS methods in health economics are relatively fewer in number, but the literature is growing. Recently, Potoglou et al. compared DCE and BWS head-to-head in their study (described above). They found that the preference weights from BWS and DCE reveal broadly similar patterns in preferences, and in the majority of cases, the preference weights are not significantly different. However, they note that this broad equivalence may not hold true in all contexts, as warned by Flynn. Further developments in the field of BWS are occurring rapidly, which may enable more widespread use of BWS in economic evaluations.

For completeness of discussion, it should be noted that further variations of the BWS approach also exist. One of them is the ‘best-worst discrete choice experiment’ in which individuals are first asked to pick the best and worst options from the attribute levels in the scenario. Once the best and worst are identified, respondents are subsequently asked to choose the best and worst among the remaining attribute levels. This process is repeated until a complete order is established. For instance, for scenarios with five attribute levels, it would take two cycles to establish a complete order. These can then be analysed using sequential multinomial models.

**Conclusion**

Economic evaluation involves valuation of health services, healthcare products, practices, policies, interventions and health states. In practice, this valuation process relies on utility-based stated preference elicitation methods that can be cardinal (SG and TTO) or ordinal. Cardinal methods are most commonly used to value health states. However, they have been criticized on the grounds of cognitive complexity, difficulty of administration, potential contamination by risk attitudes, gambling effects, risk aversion, time preference and duration effects and possible violation of the underlying EUT. As a result, ordinal preference methods have gained significant popularity in health economics.
and health services research in the last two decades. This can be attrib-
uted to reduced cognitive complexity of ordinal preference tasks, lower
degree of abstract reasoning required, reduced measurement error, ease
of administration and ability to use both health and non-health out-
comes. Furthermore, ordinal methods overcome some of the potential
biases and limitations of the traditional cardinal approaches.

This article has discussed two common ordinal preference elicitation
methods, including DCE and ranking methods. These methods have
been used to evaluate utilities of health and non-health outcomes, to
quantify trade-offs between attributes, to estimate willingness-to-pay
and to predict expected uptake of policies and products. The usefulness
of ordinal methods for valuation purposes is facilitated by the ability
to estimate cardinal values from ordinal information obtained from
DCEs and ranking methods. Ordinal valuation methods can be seen as
standalone studies or they can form part of CEA, CUA or cost–benefit
analysis.

This article highlights that, while ordinal methods are very attractive,
they are not without methodological challenge. Studies have found that
during DCEs, respondents employ cognitive heuristics (or shortcuts) to
reduce the cognitive burden of the task and often ignore attributes that
are of less interest. Moreover, respondents may not understand the
ordinal choice task as intended, or provide irrational responses\(^{96,97}\), or
make assumptions about unspecified attributes beyond the observed
choice scenarios. Finally, some authors have challenged the underlying
theoretical assumptions of DCEs when used in practice. These chal-
lenges may have implications on the validity of the estimates obtained
from DCEs.

As an alternative, ranking exercises are also becoming popular for
valuation purposes. The most common ranking exercise in health eco-
nomics is the BWS, which identifies the attribute levels that are furthest
apart on the latent utility scale. BWS has the advantage of being able
to compare attributes (and not levels only, as in DCE). BWS is relative-
ly new to the discipline and further developments in the field are occur-
ing rapidly.

This article provides a detailed discussion of the methods and chal-
lenges of using DCEs and ranking exercises for valuation purposes.
These methods have gained significant popularity since their introduc-
tion to health economics in 1990. However, this is a growing area of
research and further developments can be expected in years to come.
References


51 Schoemaker PJH. The expected utility model: its variants, purposes, evidence and limitations. 
J Econ Lit 1982;529–63.
52 Brazier J, Rowen D, Yang Y et al. Using rank and discrete choice data to estimate health state 
utility values on the QALY scale. 2009.
53 Bleichrodt H. A new explanation for the difference between time trade-off utilities and stand-
54 Flynn TN, Louviere JJ, Peters TJ et al. Best-worst scaling: what it can do for health care re-
search and how to do it. J Health Econ 2007;26:171–89.
55 Salomon JA. Reconsidering the use of rankings in the valuation of health states: a model for 
estimating cardinal values from ordinal data. Popul Health Metr 2003;1:12.
56 Moshkovich HM, Mechitov AI, Olson DL. Ordinal judgments in multiattribute decision 
57 Coast J, Flynn T, Sutton E et al. Investigating Choice Experiments for Preferences of Older 
People (ICEPOP): evaluative spaces in health economics. J Health Serv Res Policy 
58 Bridges JFP, Hauber AB, Marshall D et al. Conjoint analysis applications in health—a check-
list: a report of the ISPOR Good Research Practices for Conjoint Analysis Task Force. Value 
Health 2011.
60 Fanshel S, Bush JW. A health-status index and its application to health-services outcomes. 
61 Bradley RA, Terry ME. Rank analysis of incomplete block designs: I. The method of paired 
63 Louviere JJ, Woodworth G. Design and analysis of simulated consumer choice or allocation 
64 McFadden D. Conditional Logit Analysis of Qualitative Choice Behavior. Berkeley, 
65 Propper C. Contingent valuation of time spent on NHS waiting lists. Econ J 
66 Ryan M, Hughes J. Using conjoint analysis to assess women’s preferences for miscarriage 
67 Kind P. Applying paired comparisons models to EQ-5D valuations-deriving TTO utilities 
from ordinal preference data. EQ-5D Concepts and Methods: A Developmental History. The 
69 Carson RT, Louviere JJ. A common nomenclature for stated preference elicitation 
70 Ryan M, Farrar S. Using conjoint analysis to elicit preferences for health care. BMJ 
2000;320:1530.
71 Ryan M, Gerard K. Using discrete choice experiments to value health care programmes: 
current practice and future research reflections. Appl Health Econ Health Policy 
2003;2:55–64.
72 Scott A. Eliciting GPs’ preferences for pecuniary and non-pecuniary job characteristics. 
73 Bansback N, Brazier J, Tsuchiya A et al. Using a discrete choice experiment to estimate soci-
estal health state utility values. HEDS Discussion Paper 10/03. 2010,1–34.
74 Ryan M, Skåtun D, Major K. Using discrete choice experiments to go beyond clinical out-
comes when evaluating clinical practice. Using Discrete Choice Experiments to Value Health 
75 Hole AR. A comparison of approaches to estimating confidence intervals for willingness to 
76 Lancsar E, Wildman J, Donaldson C et al. Deriving distributional weights for QALYs 
through discrete choice experiments. J Health Econ 2011;30:466–78.


