Pharmacoeconomics 2008; 26 (8): 661-677
1170-7690/08/0008-0661/$48.00/0

Conducting Discrete Choice Experiments to Inform Healthcare Decision Making

A User's Guide

Emily Lancsar[1] and *Jordan Louviere*[2]

1 Business School (Economics) and Institute of Health and Society, University of Newcastle upon Tyne, Newcastle upon Tyne, UK
2 Centre for the Study of Choice, University of Technology, Sydney, New South Wales, Australia

Contents

Abstract

Discrete choice experiments (DCEs) are regularly used in health economics to elicit preferences for healthcare products and programmes. There is growing recognition that DCEs can provide more than information on preferences and, in particular, they have the potential to contribute more directly to outcome measurement for use in economic evaluation. Almost uniquely, DCEs could potentially contribute to outcome measurement for use in both cost-benefit and cost-utility analysis.

Within this expanding remit, our intention is to provide a resource for current practitioners as well as those considering undertaking a DCE, using DCE results in a policy/commercial context, or reviewing a DCE. We present the fundamental

principles and theory underlying DCEs. To aid in undertaking and assessing the quality of DCEs, we discuss the process of carrying out a choice study and have developed a checklist covering conceptualizing the choice process, selecting attributes and levels, experimental design, questionnaire design, pilot testing, sampling and sample size, data collection, coding of data, econometric analysis, validity, interpretation and welfare and policy analysis.

In this fast-moving area, a number of issues remain on the research frontier. We therefore outline potentially fruitful areas for future research associated both with DCEs in general, and with health applications specifically, paying attention to how the results of DCEs can be used in economic evaluation. We also discuss emerging research trends.

We conclude that if appropriately designed, implemented, analysed and interpreted, DCEs offer several advantages in the health sector, the most important of which is that they provide rich data sources for economic evaluation and decision making, allowing investigation of many types of questions, some of which otherwise would be intractable analytically. Thus, they offer viable alternatives and complements to existing methods of valuation and preference elicitation.

Given exponential increases in viable health technologies, the perennial economic problem of limited resources and unlimited claims on resources is particularly relevant in the health sector. Scarcity, coupled with the need to make choices between competing claims on resources, has focused attention on economic evaluation, ranging from evaluation of individual pharmaceuticals to evaluation of appropriate forms of healthcare financing and service delivery, all of which require valuation of healthcare and/or health outcomes. In parallel, governments and other funders are increasingly interested in public and patient preferences to inform clinical/policy decision making and improve adherence with clinical/public health programmes. In planning appropriate levels of healthcare provision, information on expected demand is also crucial.

The usual source of information on the value attached to, and preferences and demand for, goods and services is market or revealed preference (RP) data. However, RP data are scarce in health because of (i) public/private insurance, which means consumers rarely face market prices; (ii) agency relationships common in health between patients and doctors mean it is unlikely that observed consumption is based solely on patient preferences; and (iii) existence of interventions not yet in the market for which (by definition) market data do not exist.[1] This suggests a role for stated preferences (SP), or what individuals say they would do rather than what they are observed to do. SP methods commonly used in the health sector to investigate preferences and to value health outcomes include standard gamble, time trade-off, person trade-off and contingent valuation.[2] More recently, discrete choice experiments (DCEs) have been added to this list.

DCEs involve generation and analysis of choice data, and creation of hypothetical markets that can be constructed to suit relevant research questions. Thus, DCEs can mimic existing markets or elicit preferences and values for goods/services for which markets do not exist. DCEs offer several advantages in the health sector, the most important of which is that they provide rich data sources for economic evaluation and decision making, allowing investigation of many types of questions, some of which would otherwise be intractable analytically.

DCEs typically are implemented in surveys comprising several choice sets, each containing hypothetical options between which respondents choose. Each option is described by a set of attributes, and each attribute takes one of several levels. Levels describe ranges over which attributes vary across options. For example, when choosing between GPs, a key attribute might be travel time, with levels such as 5, 15 or 60 minutes. Respondents make decisions about quality- or price-differentiated versions of a good/service in a way that often requires them to

make trade-offs between attributes. The resulting choices are analysed to estimate the contribution of the attributes/levels to overall utility.

DCEs evolved out of research on axiomatic conjoint measurement[3,4] and information integration theory[5] in psychology, random utility theory-based discrete choice models in economics,[6] discrete multivariate statistical models for contingency (crosstab) tables,[7] and the optimal design of statistical experiments.[8] DCEs were pioneered in marketing by Louviere and Woodworth,[9] but quickly spread into other fields including applied economics, particularly transport[10] and environmental economics.[11]

Since the first health application in the early 1990s,[12] the number of studies using DCEs has grown rapidly (see Ryan and Gerard[13] and Ryan et al.[14] for reviews of the literature and method). Despite being popular, DCE health applications have been criticized;[15,16] and while much of the critique by Bryan and Dolan[15] was fair, Lancsar and Donaldson[17] noted that their critique largely applied to early DCE health applications, and was not a critique or invalidation of DCEs *per se*. This raises two important points: (i) to some extent best practice in DCEs has been a moving target; and (ii) it is unwise to apply DCEs without thoroughly understanding the theory, the method and how to interpret the results. This highlights a need for guidance on proper design, application, estimation and interpretation of DCEs.

Thus, the objectives of this article are to provide (i) an overview of basic DCE principles; (ii) guidance on key factors to consider in undertaking and assessing the quality of DCE applications, including a detailed checklist; and (iii) an outline of the research frontier. Our intention is to provide a resource for current practitioners and those considering undertaking a DCE, using DCE results in a policy/commercial context or reviewing DCEs.

1. Discrete Choice Experiments (DCEs) and Economic Evaluation

Despite a longer tradition of cost-benefit analysis (CBA) in economics, the dominant forms of evaluation in health economics have been cost-effectiveness analysis (CEA) and cost-utility analysis (CUA). All three approaches combine benefits with the resource use required to achieve these benefits. A key difference is the definition and scope of benefits, moving from use of intermediate uni-dimensional outcomes measured in physical units, such as change in peak flow, in CEA, to a two-dimensional unit capturing health-related quality of life (HR-QOL) and length of life, measured by QALYs,[18] in CUA, to potentially capturing all forms of benefit (including health, non-health and process benefits) using monetary valuation in CBA.

An obvious use of economic evaluation in the health sector is to evaluate pharmaceuticals and health technologies. Assessment agencies around the world primarily make decisions on value expressed as cost per QALY. A key advantage of measuring outcomes using QALYs is their generic nature, which can avoid the need for repeated valuation exercises. However, when making decisions in the health sector, consumers (and providers) may want to maximize more than QALYs.[19] It has also been noted that QALYs measure health-related utility only under specific restrictions on consumers' utility functions,[20,21] which has lead to renewed interest in CBA and valuation of benefits using willingness to pay (WTP).

So where do DCEs fit in economic analysis in health? We see DCEs contributing in two main areas: (i) eliciting preferences, quantifying trade-offs and predicting uptake to inform policy development and analysis; and (ii) measuring outcomes for inclusion in economic evaluation. Initially, applications focused on the first area, primarily eliciting patient preferences and trade-offs for features describing products or programmes in clinical settings,[22-24] and also in broader contexts such as GP's preferred remuneration packages,[25] preferred health insurance packages,[26] types of health service configurations,[27,28] and exploring time preference.[29] The initial focus was on non-health outcomes and process characteristics, with less attention paid to valuing health outcomes. More recently, recognition is growing that DCEs can provide more than preference information; for example, DCEs can be used to study the expected uptake of new policies/products[30-32] and value health outcomes.[33,34]

Almost uniquely, DCEs have the potential to provide inputs to both CBA and CUA. DCEs are

increasingly used to elicit WTP for individual characteristics of goods/services and monetary measures of benefits as a whole, which potentially could be used in CBA.[35,36] DCEs facilitate valuation of multiple options rather than evaluating a single intervention or treatment. The feasibility of using DCEs to elicit utility weights for calculation of QALYs is also being explored.[33,37,38] A possible advantage of using DCEs to elicit such weights is their grounding in utility theory. However, we are unaware of DCE-derived outcome measures being used in CBA or CUA to date; we return to the potential expanded role for DCEs in economic evaluation in section 4.

2. Theoretical Basis

DCEs represent an integration of several theoretical areas. They are consistent with Lancaster's characteristics theory of demand:[39] consumers have preferences for and derive utility from underlying attributes, rather than goods *per se*. DCEs are also consistent with welfare and consumer theory.[35,40] The two main approaches to consumer theory are preference based and choice based. The former assumes that decision makers have a preference relation over a set of possible choices that satisfies certain axioms (completeness, transitivity, monotonicity, local non-satiation, convexity and continuity), while the latter focuses on decision makers' choices, which are assumed to be consistent with the weak axiom of revealed preference (see comparison in Lancsar and Louviere[41]). The DCE approach to preference elicitation is akin to the choice-based approach to consumer theory because it explicitly assumes that choices observed in DCEs 'reveal the preferences' of individuals. Hypothetical alternatives offered in DCE surveys are constructed using experimental design theory, which is discussed in section 3.

Choices made in DCEs are analysed using random utility theory (RUT),[6,42] which posits that utility (U) for individual i conditional on choice j can be decomposed into an explainable or systematic component V_{ij} and a non-explainable or random component ε_{ij} (equation 1):

$$U_{ij} = V_{ij} + \varepsilon_{ij}\ ,\ j = 1,..., J \qquad \text{(Eq. 1)}$$

Economists view random components as being due to unobservable or unobserved attributes, unobserved preference variation, specification error and/or measurement error,[43] whereas psychologists view this component as being due to inherent variability within and between individuals.[44] The systematic component is a function of (at least) attributes of the good/service and characteristics (co-variates) of individual choosers, often modelled as shown in equation 2:

$$V_{ij} = X_{ij}'\beta + Z_i'\gamma \qquad \text{(Eq. 2)}$$

where X_{ij}' is the vector of attributes, usually including price and quality, of good j as viewed by individual i, and Z_i' is a vector of characteristics of individual i, and β and γ are vectors of coefficients to be estimated.

Utility is a latent, unobserved quantity; we observe only indicators of utility, namely choices. We assume a respondent chooses option 1 if, and only if, its utility is higher than the utility of any other option in the set of J alternatives. Assuming a joint probability distribution for ε_i, the probability P that utility is maximized by choosing option 1 is given by equation 3:

$$\begin{aligned} P(Y_i = 1) &= P(U_{i1} > U_{ij}) \\ &= P(V_{i1} + \varepsilon_{i1} > V_{ij} + \varepsilon_{ij}) \\ &= P(V_{i1} - V_{ij} > \varepsilon_{ij} - \varepsilon_{i1})\ \forall j \neq 1 \end{aligned} \qquad \text{(Eq. 3)}$$

where Y_i is a random variable denoting the choice outcome. Estimable choice models are derived by assuming a distribution for the random component. For example, if the errors are independently and identically distributed (iid) as extreme value type 1 random variates, this results in a conditional logit specification for the choice probabilities (equation 4):

$$P(Y_i = 1) = \frac{e^{\mu V_{il}}}{\sum_{j=1}^{J} e^{\mu V_{ij}}}\ ,\ j = 1,..., J \qquad \text{(Eq. 4)}$$

Using equation 2, equation 4 can be rewritten as shown in equation 5:

$$P(Y_i = 1) = \frac{e^{\mu(X_{il}'\beta + Z_i'\gamma)}}{\sum_{j=1}^{J} e^{\mu(X_{ij}'\beta + Z_i'\gamma)}}\ ,\ j = 1,..., J \qquad \text{(Eq. 5)}$$

Equations 4 and 5 have an embedded scale parameter, μ, that is inversely proportional to the variance of the error distribution, σ_ε; thus parameter estimates returned by estimation algorithms are β/σ_ε, not β.[45] μ cannot be identified in any one data source, so it is usually set to one;[11] ratios of scale parameters can be identified from two or more data sources. Such 'variance-scale ratios' account for differences in unobserved variability in the data sources, and can be specified as functions of observables.[11,44,46]

3. Undertaking a DCE

DCEs involve three main inter-related components: (i) an experimental design used to implement the choice survey and generate choice data; (ii) discrete choice analysis to estimate preferences from the choice data; and (iii) use of the resulting model to derive welfare measures and conduct other policy analyses. We discuss each in turn and summarize these issues in a checklist provided in table I.

3.1 Designing an Experiment to Generate Choice Data

3.1.1 Conceptualizing the Choice Process

Proper design and implementation of DCEs requires consideration of the choice context, nature and composition of choice sets, and framing of choice questions and instructions. DCE choice questions must be incentive compatible so as to encourage respondents to reveal true preferences.[47]

DCEs involve asking respondents to make discrete choices, in contrast with other SP methods such as conjoint analysis ranking and rating tasks. Louviere and Lancsar[48] compare these methods, and suggest reasons why traditional conjoint analysis is unlikely to be an appropriate way to elicit preferences or derive welfare measures.

Types of choice formats must be evaluated and should simulate the actual choice of interest as closely as possible. Examples include choice between pairs of alternatives, among multiple options, or binary yes/no choices. A related decision is whether the choice alternatives should be labelled (e.g. chiropractor, physiotherapy) or generic (e.g. drug A, drug B). Labelled alternatives are specified in econometric analyses with alternative specific constants (ASCs). Unless respondents must consume the good/service in practice, choice among hypothetical pairs (common in health applications) may be problematic as it implicitly assumes all respondents choose to consume the good/service,[49,50] forcing respondents to choose between two potentially unappealing alternatives, neither of which may be chosen in practice. This raises questions of how to interpret the resulting preferences because they are conditional on respondents consuming the good. Thus, allowing respondents to opt out, choose neither option, or choose status quo options should be considered, especially if an objective is to derive welfare measures.

Modelling participation/uptake is particularly relevant for investigating policies that depend on voluntary participation such as lifestyle or other population health programmes.[30] From an evaluation perspective, the comparator of interest is often a status quo treatment. If a status quo or opt out option is included, researchers must understand what this means to respondents; for example, a status quo might be a reference point for gains and losses consistent with prospect theory.[51] Status quo options can be constant for all respondents or can vary. If it varies, researchers should consider using what we call a 'report card' that asks respondents to report the attribute levels that most closely describe their particular status quo option; with the reported values then used in the model estimation (for example, see King et al.[31]).

Choices in health-related DCEs may be complex and/or unfamiliar. So it is important to consider how much experience/knowledge respondents have with the good, and how much background information and/or 'education' to provide to avoid respondents making assumptions or bringing outside (and unknown to researchers) information to the decision-making process. These issues are summarized in sections 1 and 5 of table I.

3.1.2 Defining Attributes and Levels

Attributes can be quantitative (e.g. waiting time) or qualitative (e.g. provider of care) and are generally identified from literature, qualitative research such as semi-structured interviews, and/or focus groups with samples of relevant respondents and experts (e.g. clinicians/policy makers).[52] DCEs

Table I. Checklist of factors to consider in undertaking and assessing the quality of a discrete choice experiment

1. Conceptualizing the choice process	Was a choice rather than ranking, rating task used? What type of choice was used: binary response, pairs, multiple options? Was a generic or labelled choice used? Was an opt-out, neither or status quo option included? If a forced choice was used, was a justification provided? Was the task incentive compatible?
2. Attribute selection	How were they derived and validated? Was the number of attributes appropriate? Was the coverage appropriate? What form was used: generic or alternative specific? Was price included? If so, was an appropriate payment vehicle used? Was risk included? If so, was it appropriately communicated?
3. Level selection	How were they derived and validated? Was the number of levels per attribute appropriate? Was an appropriate range used? Were the levels evenly spaced?
4. Experimental design	What type of design was used? Full factorial? Fractional factorial? If fractional, which effects are identified: main effects; main effects + higher order interactions? How were the profiles generated and allocated to choice sets? What are the properties of the design? What is the efficiency of the design? Was identification checked (e.g. is the variance-co-variance matrix block diagonal)? Was the design blocked into versions? If so, how were choice sets allocated to versions? Were the resulting properties of the versions checked? Were respondents randomly allocated to versions? How many choice sets were considered per respondent? If some profiles were implausible – how was implausibility defined and how was it addressed?
5. Questionnaire design	Was an appropriate level of background and contextual information provided? Were the task instructions appropriate? Was the medium used to communicate attribute/level information (e.g. words, pictures, multi-media) appropriate?
6. Piloting	Was coverage of attributes and levels checked? Was understanding and complexity checked? Was the length and timing checked?
7. Population/study perspective	Appropriate for research question?
8. Sample and sample size	Were inclusion/exclusion criteria explicit? Was sample size appropriate for model estimation?
9. Data collection	What recruitment method was used? How were data collected (e.g. mail, personal interview, web survey)? What was the response rate? Were incentives used to enhance response rates?
10. Coding of data	Was coding explicitly discussed? Was the coding appropriate for effects to be estimated?
11. Econometric analysis	Were the estimation methods appropriate given experimental design and type of choice response? Was the functional form of the indirect utility functions appropriate given the experimental design? Were alternative specific constants included? Were sociodemographics and other co-variates included? Was goodness of fit considered?
12. Validity	Was internal or external validity investigated? Were answers for any respondents deleted and if so on what basis?
13. Interpretation	Was the interpretation appropriate given coding of data? Were results in line with *a priori* expectations? Were relative attribute effects compared using a common and comparable metric?
14. Welfare and policy analysis	Was willingness to pay estimated using welfare theoretic compensating variation? Was probability analysis undertaken? Were marginal rates of substitution calculated?

may not include every attribute important to every respondent, but it is important to capture attributes salient to the majority to avoid respondents making inferences about omitted attributes. Lancsar and Louviere[41] discuss methods to use in pilot tests to identify whether respondents consider omitted attributes. Another consideration is whether attributes should be generic (same levels for all alternatives) or alternative specific (some attributes and/or levels differ across alternatives).

Levels should be plausible and policy/clinically relevant, although DCEs can include currently unavailable but possible alternatives (e.g. 'new horizon medications') by stretching level ranges. Indeed, a sufficiently wide range of levels should be used to avoid respondents ignoring attributes because of little difference in levels. Level range is particularly important for the price attribute if it is to be used to calculate implicit prices of other attributes using marginal rates of substitution (MRS). For example, Slothuus Skjoldborg and Gyrd-Hansen[53] found that changing the price vector changed parameter estimates and MRS; however, they noted that changing the price vector compromised the experimental design, perhaps biasing results. In contrast, Hanley et al.[54] used an experiment to study the impact of changing the price vector and found no significant impact on estimates after controlling for differences in variability between samples (variance-scale ratios noted in section 2). The payment vehicle (and duration) should be chosen to match the type of good and setting, which is well known in contingent valuation.[55] Special attention is required to properly describe risk attributes (e.g. risk of morbidity or mortality associated with different health states), as evidence suggests that people may have difficulty interpreting probabilities.[56]

Types of attribute effects to be estimated should also be considered; for example, two-level attributes only allow estimation of a linear effect, yet attributes often exhibit non-linear effects. Evenly-spaced attribute levels can be useful for interpreting the estimated effects of numerical attributes.

Specification of suitable numbers of attributes and levels is context specific; however, DCEs in health have varied as many as 12 attributes.[57] In some settings, achieving clinical relevance can require detailed attributes and levels, thereby increasing complexity of the design. Researchers may inadvertently cause omitted variable bias by excluding key attributes, and this needs to be weighed against task complexity due to too many attributes, which increases response variability. Typically, rigorous and iterative piloting is used to get the balance right. These issues are summarized in sections 2, 3 and 5 of table I.

3.1.3 Creating Experimental Design

The DCE data generation process rests heavily on an experimental design used to construct attribute combinations and choice sets. The design produces the estimation matrix, and respondents provide the dependent variable (choices) and co-variates such as sociodemographics. Thus, unlike RP data, properties of design/estimation matrices are fixed and known in advance. Thus, it behoves researchers to use optimal designs.

An experimental design is a sample from all possible combinations of attribute levels used to construct choice alternatives (or 'profiles') and assign them to choice sets. A complete census of all attribute level combinations is a 'full factorial' design. For example, if there are A attributes and all have L levels, the full factorial is L^A. A full factorial allows estimation of all main effects (effect of each attribute) and interaction effects (effect of interaction between two or more attributes) independently of one another. The number of profiles in full factorials is therefore predetermined by the dimensions of the attributes and levels. However, a full factorial often is too large to be used in practice, and therefore a 'fractional factorial' is typically used. A fractional factorial is a sample from the full factorial selected such that all effects of interest can be estimated (at a minimum, the main effects, but also as many higher-order interaction effects as possible).

The experimental design influences the types of indirect utility functions (IUFs) that can be estimated from choices, so IUF functional forms should be considered *a priori*. The design should allow estimation of the most general specification possible given constraints. Small fractional factorial designs known as orthogonal main effects plans (OMEPs), implying strictly additive IUFs, typically have been used in health. This may be convenient but is rarely likely to be correct. If IUFs are not strictly additive,

main effects are likely to be biased. Lusk and Norwood[58] suggest this is not the case, although their simulation study used parameter values for nonlinear terms and interactions that were so small that their IUFs were close to additive.[59] More work is needed in this area. In the meantime, larger fractional designs that allow estimation of (at least) all two-way interactions minimize potential for bias in main effects and allow tests of whether additive IUFs are correct. Generally, we recommend avoiding small fractional designs (i.e. designs only allowing estimation of main effects) when possible and instead recommend implementing the largest possible design given constraints such as research budgets and/or more subjective considerations of numbers of attributes and task complexity.

Full factorials may be more feasible than many researchers think, particularly because they can be blocked into different versions, with respondents randomly assigned to versions. This provides more design points without increasing numbers of choice sets for any one respondent. For example, if there are five attributes, three with four levels and two with two levels, the full factorial produces 256 ($4^3 \times 2^2$) combinations. This can be blocked into 16 versions of 16 choice sets, with respondents randomly assigned to a version. Fractional factorial designs also can be blocked into versions.[31] Typically, versions are created by randomly assigning choice sets from the design to versions without replacement; it may be possible to improve on this assignment, but as the number of attributes and levels increase, it becomes difficult to avoid correlated attributes within versions. Nonetheless, one typically can ensure that all levels of each attribute appear at least once in each block. If blocks are used, a version variable should be included in the estimation to control for version effects.

Designs can be obtained from catalogues, created using software or by hand. However they are created, their properties must be examined. Two key statistical issues in design construction are identification and efficiency. Identification determines the effects that can be estimated independently, which determines the possible IUF specifications. Independence of effects is determined by the structure of the inverse of the variance-co-variance matrix of the parameter estimates, denoted C^{-1} (where C is known as the Fisher Information Matrix). Effects are independent if C^{-1} is block diagonal. Efficiency refers to the precision with which effects are estimated; more efficient designs give more precise parameter estimates for a given sample size. For example, a design that is 50% efficient, effectively 'throws away' half the sample observations. The efficiency of a particular design typically is measured relative to that of an optimally efficient design for the particular problem of interest. A widely used efficiency criterion is D-efficiency (equation 6):

$$\text{D-Efficiency} = [\det(C)/\det(C_{opt})]^{1/\rho} \quad \text{(Eq. 6)}$$

where ρ is the number of parameters to be estimated in the model, C is defined above and C_{opt} is the largest value of the C matrix and det refers to the determinant. Street et al.[60] noted that many designs in the literature on DCEs exhibit identification problems, such that one or more effects estimated in fact were perfectly confounded with one or more other effects. Design of DCEs is entirely under the control of the researcher, so such identification problems should not occur.

Street and Burgess[61] developed theory to produce optimally or near optimally efficient designs for conditional logit models with strictly additive IUFs. Their designs create generic main effects DCEs for any choice set size for any number of attributes with any number of levels. They also provide theory to construct DCEs for main effects plus interactions if all attributes have two levels. While not yet available, research is in progress on optimally efficient designs for experiments with main effects and interactions for more than two levels. Unfortunately, except in very restrictive circumstances, optimally efficient designs for alternative-specific (labelled) DCEs are not yet available. For the latter problems, L^{MA} designs are available, where L is number of levels, A is number of attributes and M is number of choice sets.[38,62]

As noted earlier, it is often appropriate to include constant alternatives such as 'none of these' or status quo in choice sets. Such options can reduce design efficiency, but this is typically outweighed by better congruency with consumer theory and grounding in reality. It is worth noting that optimally

efficient designs for generic choices are also optimal when a 'none of these' option is included.[61]

In health, profiles are often obtained from statistical software packages such as SPEED (Stated Preference Experiment Editor and Designer), SPSS and SAS, and choice sets constructed by randomly selecting one profile and pairing it with all others. This is not only an inefficient way to construct DCEs,[61] it also can lead to identification problems. Also, some software options produce efficient designs, but these designs may not be block diagonal, resulting in parameter estimates being at least somewhat confounded with model intercept(s) and/or some or all other attributes. Hence, the resulting estimates are not independent.[61]

Huber and Zwerina[63] propose what they consider to be desirable design criteria: (i) orthogonality, i.e. attribute levels appear in choice sets with equal frequency with each level of each other attribute; (ii) level balance, i.e. levels of each attribute appear equally often; (iii) minimum overlap of levels for each attribute in each choice, and (iv) utility balance, i.e. options in each choice set have similar probabilities of being chosen. Street and Burgess[61] note that satisfying these properties does not guarantee an optimal design, and some designs that satisfy these criteria may not be identified. For example, level balance is unnecessary for an optimal design, and while minimal overlap is associated with optimal generic main effects designs, it precludes estimation of interactions. Viney et al.[38] showed that utility balance can increase the variance of the error component, which, as highlighted in section 2, can impact parameter estimates. Furthermore, if all options in each set are approximately equal in utility, there would be no reliable statistical information for model estimation. Some designs can also lead to choice sets with identical profiles, which in generic designs is a design flaw that should be corrected.

Health applications generally have used small numbers of choice sets, often eight.[13] The appropriate number of choice sets is context specific, but there is evidence that respondents can cope with more than previously considered. For example, 32 choice sets per respondent have been reported in the broader literature,[64,65] with as many as 28 used in health applications.[31] Few studies have compared responses from individuals administered small versus large numbers of choice sets.[66-68] Evidence suggests as numbers of attributes and/or choice options and/or choice sets and/or attribute differences increase, task complexity increases, which can increase unobserved variability.[66] New evidence suggests that these factors increase unobserved variability at approximately a logarithmic rate.[67] Thus, decisions about these factors should be based on realistically simulating the market of interest (i.e. as complex as the market, but no more so), and explored in iterative pilot tests.

An important issue is the possibility of and methods to handle implausible attribute combinations. That is, some minimum level of attribute A may need to be present before attribute B becomes relevant, or a level of attribute A may make no sense if combined with a level of attribute B. For example, an asthma medication that enables patients with asthma to participate in all strenuous/sporting activity but does not allow participation in daily activities makes little sense. Possible solutions involve nesting attributes (e.g. high ability of sporting activity nested with high ability to undertake daily activities), applying constraints between levels when creating designs and/or randomly replacing implausible profiles with plausible profiles. The first strategy may mean that effects of nested attributes cannot be separated; the last two strategies involve trade-offs between increased realism and reduced statistical efficiency, making it imperative to check the resulting design properties. When considering implausible combinations, implausibility should be defined from respondents' perspective rather than clinically, and thereby requires pilot testing. Key issues discussed above regarding the creation of an experimental design are summarized in section 4 of table I.

3.1.4 Pilot Tests

As with all primary data-collection methods, iterative face-to-face pilot testing is needed to guide development and testing of DCE surveys. This includes testing respondent understanding of choice contexts, generation and testing of appropriateness and understanding of attributes/levels, task complexity, length, timing and likely response rates. The importance we place on pilot testing is noted by the fact that we have discussed the need for piloting in

the various stages of developing a DCE. We summarize these issues in section 6 of table I.

3.1.5 Sample

Sampling requires consideration of the population to whom the results will be generalized, opportunity costs regarding how programmes are funded and relevant perspective (*ex ante* or *ex post*). Each has implications for relevant samples, or whose preferences to elicit, such as patients, care providers, tax payers/general public, policy makers/insurers. For example, if the good/service is to be paid for by private finance, the opportunity cost is the alternative use of individual income; this suggests that the population of interest is individual patients/users.[69] If, instead, the product/programme is to be paid out of taxes, often the case in economic evaluation in health, then the opportunity cost is the alternative use of these taxes; here the population of interest is taxpayers or the general population. Additionally, if interested in *ex post* preferences, users of the good/service are appropriate. Regardless of whose preferences are elicited, inclusion and exclusion criteria should be made explicit.

Sample size should be chosen to allow estimation of reliable models, subject to research budget and other constraints. Calculation of optimal sample sizes for estimating non-linear discrete choice models from DCE data is complicated as it depends on the true values of the unknown parameters estimated in choice models. It is also related to experimental design since the number of observations depends on the number of choice sets per respondent and number of respondents in the sample. A useful discussion of sampling for choice models (primarily for RP data) is provided in Ben-Akiva and Lerman.[43] Louviere et al.[62] provide potentially useful sample size calculations for DCEs. Specifically, if all respondents receive the same design, the minimum sample size is related to the precision of the empirical choice proportions associated with each alternative in each choice set. Otherwise, sample size is dictated by numbers of choice sets and numbers of versions. Our empirical experience is that one rarely requires more than 20 respondents per version to estimate reliable models, but undertaking significant *post hoc* analysis to identify and estimate co-variate effects invariably requires larger sample sizes. See sections 7 and 8 of table I.

3.1.6 Data Collection

Methods of data collection are well documented,[70] but we note that self-complete postal DCE surveys are common in health, often resulting in low response rates.[13] Face-to-face interviews are also used, but mini-labs in which respondents complete DCEs in central locations, or online surveys, may be more cost effective.[70] Mode of data collection is influenced by study objectives; different modes may involve different biases, which are well documented elsewhere.[70] Issues to consider are summarized in section 9 of table I.

3.2 Discrete Choice Analysis

3.2.1 Coding

Coding of explanatory variables is important for analysis and interpretation of results, particularly ASCs and interactions. Typically, effects coding or dummy variable coding are used, particularly for qualitative attributes. Mean-centering numerical attributes can be useful when specifying non-linear effects for numerical attributes such as quadratic or cubic effects. Effects codes and mean-centering avoid correlations with the ASCs/intercepts, allowing the ASCs/intercepts to be interpreted as reflecting aggregate shares of choices and minimizing collinearity in estimation matrices used to estimate interactions. Several studies in health economics have used effects codes;[30,31,35] the importance of using these codes is highlighted by Bech and Gyrd-Hansen.[71] It is worth noting that the estimate of the omitted level of an effects-coded attribute is simply minus one times the sum of the estimated levels. Table II shows an example of effects coding for a four-level attribute. As can be seen in the table, this is very similar to dummy coding in that only L − 1 levels (three in this case) are coded, with the omitted Lth level on each effects coded variable coded −1

Table II. Effects coding (Fx) for a four-level attribute

Levels	Fx1	Fx2	Fx3
0	1	0	0
1	0	1	0
2	0	0	1
3	−1	−1	−1

rather than 0. Coding is considered in section 10 of table I.

3.2.2 Forms of Choice Models

The form of the estimated IUF depends on (i) the experimental design and whether interaction effects are identified and/or alternatives are labelled; and (ii) the type of choice modelled (binary choices imply binary models; multiple choices imply multinomial models). For example, the conditional logit model (CLM) of equation 4 is a fixed effects logit model that provides a closed form solution for the choice probabilities and is easily estimated. A key property of the CLM associated with the iid assumption is the independence of irrelevant alternatives (IIA) that implies proportional substitutability across alternatives.[45] Whether IIA holds is an empirical question, tests for which are outlined in the article by Train.[45]

Different choice models arise from different assumptions about distributions and properties of error components and about variance-co-variance matrices of preference parameters. For example, the nested logit model relaxes IIA by allowing violations of IIA between nests, while requiring IIA to hold within nests. Other models that relax IIA include multinomial probit (for a health-related example, see Ryan et al.[37]) and mixed logit (MIXL) [health-related examples include Johnson et al.,[22] Hall et al.,[57] Kjaer and Gyrd-Hansen[72]] to name two.

McFadden and Train[73] show that any random utility model can be approximated by an MIXL. MIXL has more flexible substitution patterns and can accommodate the panel nature of DCE data by allowing correlation within subjects over repeated choices. It also allows for preference heterogeneity across individuals by allowing parameters to vary randomly across individuals. This is achieved by including a respondent-specific stochastic component (β_i): $\beta_i = \bar{\beta} + \mu_i$, where $\bar{\beta}$ is the mean parameter vector for the population and μ_i is the individual specific deviation from the mean. One must specify a distribution for each β_i and estimate the parameters of that distribution (i.e. mean and standard deviation). MIXL does not have a closed form solution, requiring simulated maximum likelihood estimation (or hierarchical Bayes for the Bayesian versions). Other models that relax IIA include latent class models[74] and heteroscedastic error variance models.[66,75,76] While highly flexible, a potential problem with these models is that it is unlikely that error variances are constant within or between individuals,[44,64,65,77] in which case model parameters are confounded with the unobserved distribution of error variances.

Regardless of the type of choice model estimated, the functional forms of individual variables should be informed by economic theory whenever possible. In addition, we recommend estimating a model in the most disaggregated form by including parameter estimates for L − 1 attribute levels, then graphing these estimates against the levels of each attribute to visualize implied functional forms. This allows recoding and re-estimation of more parsimonious models using the implied specification. For example, if a graph suggests that the estimated utilities increase at a decreasing rate with the levels of a numerical attribute, a quadratic or logarithmic specification may be appropriate.[78] For labelled DCEs with J alternatives, ASCs for J − 1 alternatives can be included that represent the underlying preference for each alternative when attributes are effects coded and set to zero. Naturally, specifications with interactions, sociodemographic variables and co-variates should be estimated as appropriate.

Log likelihood and pseudo R-squared values can inform goodness of fit of estimated models. Model selection is informed by (i) economic and behavioural theory, and (ii) statistical considerations such as likelihood ratio tests for nested models and the Akaike information criteria (AIC) and the Bayesian information criteria (BIC) for non-nested models. Issues to consider in undertaking or reviewing econometric analysis of a DCE are summarized in section 11 of table I.

3.2.3 Validity

Validity of DCEs is relatively well established in the broader literature,[62] with comparisons to RP data in marketing, environmental and transportation economics.[11,62,79] There have been relatively few tests of external validity in health, perhaps due to limited RP data, although Mark and Swait[80] found evidence of external validity in prescribing decisions for alcoholism medication. Instead, the focus

has been on internal validity, usually limited to checking if signs of estimated parameters are consistent with *a priori* expectations; some researchers have tested if results conform with the axioms of consumer theory (e.g. completeness, monotonicity and transitivity[81-83]). Similarly, researchers have also studied 'rationality' of choices, defining 'irrational' responses by failure of non-satiation or lexicographic preferences (the latter, in fact, are not irrational), using tests to exclude 'irrational' individuals from analysis. Lancsar and Louviere[41] discussed several problems in testing 'rationality', including the fact that apparent 'irrationality' can be due to (i) shortcomings in design and implementation of DCEs; (ii) respondent learning about their preferences or tasks; (iii) 'irrationality' tests not being conclusive; (iv) use of fractional factorials, which cannot identify unique decision rules. They also provided evidence that RUT can cope with such preferences. Deleting respondents may omit valid preferences leading to bias and lower statistical efficiency. Indeed, internal validity is broader than econometric testing; for example, well designed and implemented studies, that are consistent with the previous discussion and the checklist provided in table I of issues to consider at each stage of undertaking/reviewing a DCE, give more confidence in results. Validity is considered in section 12 of table I.

3.3 Interpretation, Derivation of Welfare Measures and Other Policy Analysis

Once a preference model (the IUF) is estimated, it can be used in policy analyses in various ways, such as comparing the relative importance of product/programme attributes. For example, when choosing diagnostic tests, is test accuracy relatively more important to patients than time spent waiting for results? Many studies measure the relative impact of attributes by comparing size and significance of estimated attribute parameters. Unfortunately these parameters can not be directly compared, because attribute impacts and the positions of each attribute level on the underlying utility scale are confounded (i.e. distances between utilities associated with attribute levels need not be the same for each attribute).[78] To measure relative attribute impacts, one needs to measure each on a common, comparable, scale. See section 13 of table I. Lancsar et al.[78] discussed five ways to compare relative attribute impact, some of which we discuss below.

3.3.1 Predicted Probability Analysis

The probability that respondents will choose each alternative in a choice set is calculated using equation 4, which also allows comparison of the impact of each attribute in a common metric.[30,31,78] In the case of non-closed form models, the choice probabilities need to be simulated to approximate the integration over choice situations/respondents, but otherwise, the process is the same. Predicted probabilities are also used to evaluate expected market shares in marketing applications, and an obvious analogue in health is predicting uptake or choice shares for the sample that provided choices. To predict beyond the sample requires recalibration of DCE results, which is appropriate when market data are available.

3.3.2 Marginal Rates of Substitution

DCEs allow estimation of trade-offs that respondents make between attributes, or their MRS.[23,25,84] Following standard consumer theory, MRS is calculated by partially differentiating the IUF, equation 2, with respect to the first attribute and with respect to the second attribute, and calculating their ratio (equation 7):

$$MRS_{X_1,X_2} = \frac{\partial V / \partial X_1}{\partial V / \partial X_2} \quad \text{(Eq. 7)}$$

where V is an IUF and X_1, X_2 are attributes of the good/service and ∂ is the partial derivative. The numerator (denominator) is interpreted as the marginal utility of attribute 1 (2). If price is the numeraire, the denominator denotes the marginal disutility of price, and we term the calculation the 'implicit price' of each attribute. If the IUF is linearly additive, equation 7 equals the ratio of the estimated attribute parameters. MRS for non-linear utility functions can be used to investigate attribute impact, but the calculation is more complex as explained by Lancsar et al.[78]

3.3.3 Welfare Measures to Value Health and Healthcare

DCEs are flexible, which is an advantage for welfare measurement because the value of an entire

good/service and different configurations of goods/services can be estimated. The method of calculating Hicksian compensating variation (CV) in discrete choice random utility models in general[85] was recently introduced to health economics to calculate welfare measures in the context of DCEs.[1,35] The CV method can calculate measures of welfare gain, or WTP, for entire products/programmes, and can measure the relative impacts of each attribute in a common monetary metric as WTP or accept compensation for changes in a given attribute. For a conditional logit model, both forms of welfare measures are calculated using the utility estimates and attribute levels in the following expression (equation 8):

$$CV = -\frac{1}{\lambda}\left[\ln \sum_{j=1}^{J} e^{V_j^0} - \ln \sum_{j=1}^{J} e^{V_j^1}\right] \quad \text{(Eq. 8)}$$

where J is the number of options in the choice set; e is the exponential; λ is the marginal utility of income; and V_j^0 and V_j^1 are the value of the IUF for each choice option j before and after the policy change, respectively.

Hicksian CV basically values a change in expected utility due to a change in the attribute(s), by weighting this change by the marginal utility of income. It takes account of the uncertainty in the choice model about which alternative respondents will choose and/or whether respondents substitute among alternatives following a change in the desirability of one or more alternatives. Again, for non-closed form models, the CV needs to be simulated.[32] Equation 8 also can be used to calculate the CV using non-monetary metrics; for example Baker et al.[86] calculated WTP in terms of QALYs for a change in health state using the marginal utility of a QALY as the numeraire.

The product of the sum of MRS and the change in the attributes of interest has been used in health economics to calculate WTP for goods/services. However, as Lancsar and Savage[35] noted, that approach is generally inappropriate for welfare measurement and instead the theoretically consistent method in equation 8 should be used. Both MRS and WTP are random variables, so the uncertainty or variance in the resulting values can be captured by estimating confidence intervals. Risa Hole[87] provided a useful review and comparison of methods available to calculate confidence intervals. Issues to check regarding interpretation of DCE results and welfare and policy analysis are included in sections 13 and 14 of table I.

Of course, DCEs have potential limitations. As the forgoing suggests, designing, undertaking and interpreting DCEs can be a time-consuming and involved process. Thus, it is important to consider before commencing a study whether a DCE in fact is the most suitable method for the research question. DCEs can be cognitively demanding for respondents. Generalizability of results may be an issue in economic evaluation depending on how the DCE is designed and administered.[15,50] A new DCE may be required for each research question, although because of the flexibility of DCEs several versions of a programme or treatment can be valued within a single study. We return to some of these issues in the next section.

4. The Research Frontier

A number of issues remain on the research frontier associated with the DCE approach in general, and health applications specifically. For example, there is scope to move beyond simplistic and *ad hoc* use of qualitative methods in developing DCEs to (iteratively) applying more sophisticated qualitative tools before, alongside and after quantitative data collection. Ideally, what is required is theory or at least a systematic approach to qualitative research (including pilot testing) for developing and testing DCEs. Progress in this area is exemplified by Coast and Horrocks,[52] but qualitative methods remain underutilized.

Challenges remain in developing optimal design theory for alternative specific or labelled choices as well as choice sets with individual-specific status quo options. Also, as more complex and flexible IUFs and choice models are used, optimal design theory is needed to support such specifications. Likewise, larger designs naturally lead to blocking choice sets into versions, requiring statistical guidelines for not only how to construct optimally efficient DCEs, but also how to optimize allocation of resulting choice sets into blocks. As DCEs become more complex, we need to better understand the relationships between design efficiency and respon-

dent efficiency (influenced in part by cognitive burden), which remains under-researched despite some work.[67,88,89] Further work is also required on the nature of the potential bias arising from using designs that ignore interactions.

Sample size is another practical area warranting further research. This is likely to be challenging because estimation of sample sizes requires knowledge of the unknown parameter estimates *a priori*. A potentially fruitful research avenue would be to use pilot tests to estimate parameter values to use in sample size calculations for a more complete study.[43]

Analysis of DCEs in health has primarily focused on response means, but variances of outcome distributions and error components warrant attention.[44,64,65] Similarly, models that can deal with issues associated with differences in variance-scale ratios or, more generally, non-constant error variances are needed, as noted in section 2. Accounting for unobserved variability and preference heterogeneity is important, and the relative merits of various ways to do this need to be explored further. Theory and methods have been developed to allow one to model the choices of single individuals, which eliminates the need to make distributional assumptions about preference heterogeneity.[67]

DCEs can isolate and measure patient preferences, but decisions about healthcare treatment can involve joint decisions between patients and their doctors.[90,91] Opportunities exist to use DCEs to investigate agency relationships in health as in other sectors,[92] which complements investigating patient preferences. For example, Bartels et al.[92] used two DCEs, one for consumers and one for providers with overlap in the attributes between the two, which allowed them to investigate agency relationships in decisions regarding water heaters. This approach could prove useful for similar analysis in the health sector. Another extension includes modelling multistage choice processes, which can include quantities chosen and changes in choices over time.[44,65] Integration of more behavioural theory and incorporating contributions from various fields such as psychology should be beneficial.

Generalizability of DCE results is a key research need because it relates to factors such as time, context and geography that are often constant in single data sources.[93] Thus, we need to understand the effects of these factors, including the effects on unobserved variability. Similarly, work is needed on the extent to which DCEs can be used in benefit transfer, similar to applications in the contingent valuation literature.[94] Indeed, DCEs may be well suited to benefit transfer applications, as they are more general and flexible about the composition of goods/services than contingent valuation, which may make transferability easier.

Despite prior work on 'rationality' of DCE responses, many tests focused on axioms not strictly required for rationality.[68,95] Work focusing on axioms of transitivity and completeness and tests of the weak and strong axioms of revealed preference would be welcome. Also, evidence that respondents may not use compensatory decision making rules (IUFs) suggests that more work is needed to understand if/when alternative decision rules or heuristics are used.[81,96]

A new type of choice experiment called 'best-worst scaling' (BWS) is garnering attention in health economics and more broadly. The underlying theoretical properties of BWS were formally proven by Marley and Louviere[97] and Marley et al.[98] Thus far, two types of BWS have been used in health economics: (i) asking respondents to choose the best and worst attribute level in several single profiles, which potentially allows one to estimate the importance of each attribute and measure them on a common scale;[78,99,100] and (ii) asking respondents to choose the best and worst alternatives in each of several choice sets, allowing one to observe many more choices without increasing numbers of choice sets.[101] We expect such approaches to see increasing use in health economics.

While DCEs have mainly focused on estimating preferences for goods/services, we expect to see them used more directly in outcome measurement for use in economic evaluation. DCEs allow estimation of theoretically consistent measures of welfare gain or WTP,[35] suggesting that they can be used in CBA. McIntosh[36] proposed a framework for development of DCE-derived CBAs in health. Similarly, DCEs could also be used to inform CUA. In particular, DCEs potentially can be used to derive utility weights for calculating QALYs. This requires further research and comparison to more standard

methods such as time trade-off, standard gamble and the visual analogue scale and would require a large-scale study to investigate population values. Finally, and perhaps more interestingly, there is scope to elicit both health- and non-health-related utility that could potentially be used directly in CUA (along the lines of a 'super QALY'[102]). Considerable work would be required to test if this is feasible. The ability to measure outcomes (in utility or monetary measures) using DCEs offers additional 'tools' for the health economists' 'tool kit'.

DCEs have been embedded in randomized controlled trials (RCTs),[31,103] and are likely to be included more routinely in RCTs as they become more mainstream in economic evaluation. There may also be a role for DCEs in the expanded remit of the UK National Institute for Health and Clinical Excellence, which includes economic evaluations of population health initiatives. A key issue in such evaluations is participation/uptake, areas where DCEs can be useful. DCEs also may be useful for priority setting,[104] including serving as a component of priority-setting tools such as programme budgeting marginal analysis.[105]

5. Conclusion

This article has reviewed and discussed the application and development of DCEs in health economics more than 15 years after the first application. We have provided a checklist of issues to consider when developing or reviewing the quality of a DCE. Of course, DCEs are not a panacea, but if appropriately designed, implemented, analysed and interpreted, they offer rich sources of information to inform economic evaluation and decision making more broadly in the health sector. Thus, they offer viable alternatives and complements to existing methods of valuation and preference elicitation.

In 2002, Viney et al.[50] noted "Given the growth in the number of applications of DCEs in health in the last 5 years and the potential areas where DCEs could contribute to policy and resource allocation, it is likely that they will become a standard tool in health economics research over the next 5 years, although it may be longer before they become a standard tool for health policy." Five years on, this seems to be an accurate prediction. DCEs are now a standard health economics research tool, but they are not yet a standard health policy tool, although they are starting to be used in that way.

Acknowledgements

No sources of funding were used to assist in the preparation of this article. The authors have no conflicts of interest that are directly relevant to the content of this article.

The authors thank the anonymous referees for helpful comments.

References

1. Lancsar E. Deriving welfare measures from stated preference discrete choice modelling experiments [CHERE discussion paper no. 48]. Sydney: Centre for Health Economics Research and Evaluation, University of Technology Sydney, 2002
2. Drummond MF, Sculpher MJ, Torrance GW, et al. Methods for the economic evaluation of health care programmes. 3rd ed. Oxford: Oxford University Press, 2005
3. Krantz DH, Tversky A. Conjoint measurement analysis of composition rules in psychology. Psychology Rev 1971; 78: 151-69
4. Luce RD, Tukey JW. Simultaneous conjoint measurement: a new type of fundamental measurement. J Math Psychol 1964; 1: 1-27
5. Anderson NH. Foundations of information integration theory. New York: Academic Press, 1981
6. McFadden D. Conditional logit analysis of qualitative choice behavior. In: Zarembka P, editor. Frontiers of econometrics. New York: Academic Press, 1974: 105-42
7. Bishop YM, Fienberg SW, Holland PW. Discrete multivariate analysis. Cambridge: MIT Press, 1975
8. Box GEP, Hunter WG, Hunter JS. Statistics for experimenters. New York: Wiley, 1978
9. Louviere J, Woodworth G. Design and analysis of simulated consumer choice or allocation experiments: an approach based on aggregated data. J Mark Res 1983; 20: 350-67
10. Hensher DA, Louviere JJ. On the design and analysis of simulated choice or allocation experiments in travel choice modelling. Transport Res 1983; 890: 11-7
11. Adamowicz W, Louviere J, Williams M. Combining revealed and stated preference methods for valuing environmental amenities. J Environ Manage Econ 1994; 26 (3): 271-92
12. Propper C. Contingent valuation of time spent on NHS waiting lists. Econ J 1990; 100: 193-9
13. Ryan M, Gerard K. Using discrete choice experiments to value health care programmes: current practice and future research reflections. Applied Health Econ Health Policy 2003; 2 (1): 55-64
14. Ryan M, Gerard K, Amaya-Amaya M. Using discrete choice experiments to value health and health care. Dordrecht: Springer, 2008
15. Bryan S, Dolan P. Discrete choice experiments in health economics: for better or for worse? Eur J Health Econ 2004; 5 (3): 199-202
16. Wainright DM. More 'con' than 'joint': problems with the application of conjoint analysis to participatory healthcare decision making. Crit Public Health 2003; 13: 373-80
17. Lancsar E, Donaldson C. Discrete choice experiments in health economics: distinguishing between the method and its application [comment]. Eur J Health Econ 2005; 6 (4): 314-6
18. Torrance GW. Measurement of health state utilities for economic appraisal. J Health Econ 1986 5: 1-30
19. Nord E, Pinto JL, Richardson J, et al. Incorporating societal concerns for fairness in numerical valuations of health pro-

grammes [published erratum appears in Health Econ 1999 Sep; 8 (6): 559]. Health Econ 1999; 8 (1): 25-39
20. Bleichrodt H. QALYs and HYEs (healthy year equivalents): under what conditions are they equivalent? J Health Econ 1995; 14 (1): 17-37
21. Pliskin J, Shepard D, Weinstein W. Utility functions for life years and health status. Oper Res 1980; 28: 206-24
22. Johnson FR, Banzhaf MR, Desvousges WH. Willingness to pay for improved respiratory and cardiovascular health: a multiple-format, stated-preference approach. Health Econ 2000; 9 (4): 295-317
23. Ryan M. Using conjoint analysis to take account of patient preferences and go beyond health outcomes: an application to in vitro fertilisation. Soc Sci Med 1999; 48 (4): 535-46
24. Ryan M, Hughes J. Using conjoint analysis to assess women's preferences for miscarriage management. Health Econ 1997; 6: 261-73
25. Scott A. Eliciting GPs' preferences for pecuniary and non-pecuniary job characteristics. J Health Econ 2001; 20: 329-47
26. Chakraborty G, Ettensen R, Gaeth G. How consumers choose health insurance. J Health Care Mark 1994; 14 (1): 21-33
27. Jan S, Mooney G, Ryan M, et al. The use of conjoint analysis to elicit community preferences in public health research: a case study of hospital services in South Australia. Aust NZ J Public Health 2000; 24: 64-70
28. Morgan A, Shackley P, Pickin M, et al. Quantifying patient preferences for out-of-hours primary care. J Health Serv Res Policy 2000; 5: 214-8
29. van der Pol M, Cairns J. Estimating time preference for health using discrete choice experiments. Soc Sci Med 2001; 52: 1459-70
30. Hall J, Kenny P, King M, et al. Using stated preference discrete choice modelling to evaluate the introduction of varicella vaccination. Health Econ 2002; 11: 457-65
31. King MT, Hall J, Lancsar E, et al. Patient preferences for managing asthma: results from a discrete choice experiment. Health Econ 2007; 16 (7): 703-17
32. Lancsar EJ, Hall JP, King M, et al. Using discrete choice experiments to investigate subject preferences for preventive asthma medication. Respirology 2007; 12 (1): 127-36
33. Hakim Z, Pathak DS. Modelling the EuroQol data: a comparison of discrete choice conjoint and conditional preference modelling. Health Econ 1999; 8 (2): 103-16
34. Sculpher M, Bryan S, Fry P, et al. Patients' preferences for the management of non-metastatic prostate cancer: discrete choice experiment. BMJ 2004; 328: 382-4
35. Lancsar E, Savage E. Deriving welfare measures from discrete choice experiments: inconsistency between current methods and random utility and welfare theory. Health Econ 2004; 13 (9): 901-7
36. McIntosh E. Using discrete choice experiments within a cost-benefit analysis framework: some considerations. Pharmacoeconomics 2006; 24 (9): 855-68
37. Ryan M, Netten A, Skatun D, et al. Using discrete choice experiments to estimate a preference-based measure of outcome: an application to social care for older people. J Health Econ 2006; 25 (5): 927-44
38. Viney R, Savage E, Louviere J. Empirical investigation of experimental design properties of discrete choice experiments in health care. Health Econ 2005; 14 (4): 349-62
39. Lancaster K. A new approach to consumer theory. J Polit Econ 1966; 74: 132-57
40. Hanley N, Mourato S, Wright RE. Choice modelling approaches: a superior alternative for environmental valuation? J Econ Surv 2001; 15: 435-62
41. Lancsar E, Louviere J. Deleting 'irrational' responses from discrete choice experiments: a case of investigating or imposing preferences? Health Econ 2006; 15 (8): 797-811
42. Thurstone L. A law of comparative judgement. Psycholog Rev 1927; 34: 273-86
43. Ben-Akiva M, Lerman SJ. Discrete choice analysis: theory and applications to travel demand. Cambridge: The MIT Press, 1985
44. Adamowicz W, Bunch D, Cameron T, et al. Behavioural frontiers in choice modelling. Marketing Lett. In press
45. Train KE. Discrete choice methods with simulation. Cambridge: Cambridge University Press, 2003
46. Swait J, Louviere J. The role of the scale parameter in the estimation and comparison of multinomial logit models. J Mark Res 1993; 30: 305-14
47. Carson RT, Groves T, Machina MJ. Incentive and informational properties of preference questions. San Diego (CA): University of California, 2000
48. Louviere J, Lancsar E. Distinguishing between conjoint analysis and discrete choice experiments with implications for stated preference and welfare elicitation. Sydney (NSW): CenSoC, University of Technology, 2008
49. Ryan M, Skatun D. Modelling non-demanders in choice experiments. Health Econ 2004; 13 (4): 397-402
50. Viney R, Lancsar E, Louviere J. Discrete choice experiments to measure consumer preferences for health and healthcare. Exp Rev Pharmacoeconomics Outcomes Res 2002 August; 2 (4): 319-26
51. Kahneman D, Tversky A. Prospect theory: an analysis of decision under risk. Econometrica 1979; 47: 263-91
52. Coast J, Horrocks S. Developing attributes and levels for discrete choice experiments using qualitative methods. J Health Serv Res Pol 2007; 12 (1): 25-30
53. Slothuus Skjoldborg U, Gyrd-Hansen D. Conjoint analysis: the cost variable. An Achilles' heel? Health Econ 2003; 12 (6): 479-91
54. Hanley N, Adamowicz W, Wright RE. Price vector effects in choice experiments: an empirical test. Res Energy Econ 2005; 27: 227-34
55. Smith R. Construction of the contingent valuation market in health care: a critical assessment. Health Econ 2003; 12: 609-28
56. Peters E, Vastfjall D, Slovic P, et al. Numeracy and decision making. Psycholog Sci 2006; 17 (5): 407-13
57. Hall J, Fiebig DG, King MT, et al. What influences participation in genetic carrier testing? Results from a discrete choice experiment. J Health Econ 2006; 25 (3): 520-37
58. Lusk JL, Norwood FB. Effect of experimental design on choice-based conjoint valuation estimates. Am J Agric Econ 2005; 87 (3): 771-85
59. Louviere JJ, Wasi N. A warning about possibly misleading conclusions in Lusk and Norwood (2005) [CenSoC working paper]. Sydney (NSW): University of Technology, Sydney, 2008
60. Street DA, Burgess L, Louviere JJ. Quick and easy choice sets: constructing optimal and nearly optimal stated choice experiments. Int J Res Mark 2005; 22: 459-70
61. Street DA, Burgess L. The construction of optimal stated choice experiments: theory and methods. Hoboken (NJ): Wiley, 2007
62. Louviere JJ, Hensher DA, Swait JD. Stated choice methods analysis and application. Cambridge: Cambridge University Press, 2000
63. Huber J, Zwerina K. The importance of utility balance in efficient choice designs. J Mark Res 1996; 33: 307-17
64. Louviere J, Engle T. Confound it! That pesky little scale constant messes up our convenient assumptions. 2006 Sawtooth

Software Conference Proceedings; 2006 Mar 29-31; Delray Beach (FL). Sequim (WA): Sawtooth Software: 211-28
65. Louviere JJ, Meyer RJ. Formal choice models of informal choices: what choice modeling research can (and can't) learn from behavioural theory. Rev Mark Res 2007; 4: 3-32
66. DeShazo JR, Fermo G. Designing choice sets for stated preference methods: the effects of complexity on choice consistency. J Environ Econ Manage 2002; 44: 123-43
67. Louviere J, Islam T, Wasi N, et al. Designing discrete choice experiments: do optimal designs come at a price? J Consumer Res 2008 Aug. In press
68. San Miguel F, Ryan M, Amaya-Amaya M. Irrational stated preferences: a quantitative and qualitative investigation. Health Econ 2004; 14: 307-22
69. Shackley P, Donaldson C. Willingness to pay for publicly-financed health care: how should we use the numbers? App Econ 2000; 32: 2015-21
70. Dillman DA, Bowker DK. Mail and internet surveys: the tailored design method. New York: Wiley, 2001
71. Bech M, Gyrd-Hansen D. Effects coding in discrete choice experiments. Health Econ 2005; 14 (10): 1079-83
72. Kjaer T, Gyrd-Hansen D. Preference heterogeneity and choice of cardiac rehabilitation program: results from a discrete choice experiment. Health Policy 2008; 85: 124-32
73. McFadden D, Train KE. Mixed MNL models for discrete response. J Applied Econometrics 2000; 15: 447-70
74. Swait J, Adamowicz W. The influence of task complexity on consumer choice: a latent class model of decision strategy switching. J Con Res 2001; 28 (1): 135-48
75. Islam T, Louviere JJ, Burke PF. Modeling the effects of including/excluding attributes in choice experiment on systematic and random components. Int J Res Mark 2007; 24: 289-300
76. Swait J, Adamowicz W. Choice environment, market complexity, and consumer behavior: a theoretical and empirical approach for incorporating decision complexity into models of consumer choice. Organ Behav Human Decision Processes 2001; 86 (2): 141-67
77. Magidson J, Vermunt J. Removing the scale factor confound in multinomial logit choice models to obtain better estimates of preference. 2007 Sawtooth Software Conference; 2007 Oct 17-19; Santa Rosa (CA)
78. Lancsar E, Louviere J, Flynn T. Several methods to investigate relative attribute impact in stated preference experiments. Soc Sci Med 2007; 64 (8): 1738-53
79. Adamowicz W, Swait J, Boxall P, et al. Perceptions versus objective measures of environmental quality in combined revealed and stated preference models of environmental valuation. J Environ Manage 1997; 32: 65-84
80. Mark T, Swait J. Using stated preference and revealed preference modelling to evaluate prescribing decisions. Health Econ 2004; 13 (6): 563-73
81. Lloyd AJ. Threats to the estimation of benefit: are preference elicitation methods accurate? Health Econ 2003; 12: 393-402
82. McIntosh E, Ryan M. Using discrete choice experiments to derive welfare estimates for the provision of elective surgery: implications of discontinuous preferences. J Econ Psychol 2002; 23 (3): 367-82
83. Ryan M, San Miguel F. Revisiting the axiom of completeness in health care. Health Econ 2003; 12 (4): 295-307
84. Gyrd-Hansen D, Søgaard J. Analysing public preferences for cancer screening programmes. Health Econ 2001; 10 (7): 617-34
85. Small KA, Rosen HS. Applied welfare economics with discrete choice models. Econometrica 1981; 49 (1): 105-30
86. Baker R, Donaldson C, Lancsar E, et al. Deriving QALY weights through discrete choice experiments: challenges and preliminary results. Health Economics Study Group Meeting; 2008 Jan 9-11; Norwich
87. Risa Hole A. A comparison of approaches to estimating confidence intervals for willingness to pay measures. Health Econ 2007; 16: 827-40
88. Maddala T, Phillips KA, Reed Johnson F. An experiment on simplifying conjoint analysis designs for measuring preferences. Health Econ 2003; 12 (12): 1035-47
89. Severin V. Comparing statistical efficiency and respondent efficiency in choice experiments. Sydney (NSW): University of Sydney, 2000
90. Bryan S, Gill P, Greenfield S, et al. The myth of agency and patient choice in health care? The case of drug treatments to prevent coronary disease. Soc Sci Med 2006; 63 (10): 2698-701
91. Vick S, Scott A. Agency in health care: examining patients' preferences for attributes of the doctor-patient relationship. J Health Econ 1998; 17: 587-605
92. Bartels R, Fiebig DG, van Soest A. Consumers and experts: an econometric analysis of the demand for water heaters. Empirical Econ 2006; 31: 639-391
93. Louviere J, Street D, Carson R, et al. Dissecting the random component of utility. Marketing Lett 2002; 13 (3): 177-93
94. Brouwer R, Bateman IJ. Benefit transfer of willingness to pay estimates and functions for health-risk reductions: a cross-country study. J Health Econ 2005; 24: 591-611
95. Ryan M, Bate A. Testing the assumptions of rationality, continuity and symmetry when applyling discrete choice experiments in health care. Appl Econ Lett 2001; 8: 59-63
96. Swait J. A non-compensatory choice model incorporating attribute cutoffs. Transportation Research B 2001; 35 (10): 903-28
97. Marley A, Louviere J. Some probabilistic models of best, worst, and best-worst choices. J Math Psychol 2005; 49 (6): 464-80
98. Marley AAJ, Louviere JJ, Flynn T. Probabilistic models of set-dependent and attribute-level best-worst choice. J Math Psychol. In press
99. Flynn TN, Louviere JJ, Peters TJ, et al. Best-worst scaling: what it can do for health care research and how to do it. J Health Econ 2007; 26 (1): 171-89
100. McIntosh E, Louviere J. Separating weight and scale value: an exploration of best-attribute scaling in health economics. Health Economics Study Group Meeting; 2002 Jul 3-5; London
101. Lancsar E, Louviere J. Several methods for dealing with scale confound and efficiency in stated preference data with an empirical illustration. Health Economics Study Group Meeting; 2005 Jun 29-Jul 1; Newcastle upon Tyne
102. Anand P. QALYs and the integration of claims in health-care rationing. Health Care Anal 1999; 7: 239-53
103. Longo MF, Cohen DR, Hood K, et al. Involving patients in primary care consultations: assessing preferences using discrete choice experiments. Br J Gen Practice 2006; 56: 35-42
104. Ratcliffe J. Public preferences for the allocation of donor liver grafts for transplantation. Health Econ 2000; 9 (2): 137-48
105. Ruta D, Mitton C, Bate A, et al. Programme budgeting and marginal analysis: bridging the divide between doctors and managers. BMJ 2005; 330: 1501-3

Correspondence: *Emily Lancsar*, University of Newcastle upon Tyne, 21 Claremont Place, Newcastle upon Tyne, NE2 4AA, UK.
E-mail: Emily.Lancsar@ncl.ac.uk