Title: Conducting Discrete Choice Experiments to Inform Healthcare Decision Making: An Overview of Fundamental Principles

Running title: DCEs to Inform Healthcare Decision Making

Manuscript number: PE-1-07-00410

Article Type: Practical Application article

Emily Lancsar[1] and Jordan Louviere[2]

1. Business School (Economics) and Centre for Health Services Research, University of Newcastle upon Tyne, UK

2. Centre for the Study of Choice, University of Technology, Sydney, Australia

Name and Address for correspondence:

Emily Lancsar

University of Newcastle upon Tyne

21 Claremont Place

Newcastle upon Tyne

NE24AA

Phone: +44 0191 222 6459

Fax: +44 0191 222 6548

Email: Emily.Lancsar@ncl.ac.uk

Table of contents

Abstract

DCEs are regularly used in health economics to elicit preferences for health care products and programs. There is growing recognition that DCEs can provide more than information on preferences and in particular have the potential to contribute more directly to outcome measurement for use in economic evaluation. Almost uniquely, DCEs potentially can contribute to outcome measurement for use in both cost benefit analysis and cost utility analysis. Within this expanding remit, this paper provides: an overview of fundamental principles ofDCEs; guidance on the appropriate design, application, estimation and interpretation of DCEs, including a checklist; and an outline of the research frontier paying attention to how the results can be used in economic evaluation. 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.

1. Introduction

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 completing claims on resources has focused attention on economic evaluation, ranging from evaluation of individual pharmaceuticals, to evaluation of appropriate forms of health care financing and service delivery, all of which require valuation of health care and/or health outcomes. In parallel, governments and other funders are increasingly interested in public and patient preferences to inform clinicaVpolicy decision making and improve adherence with clinicaVpublic health programs. In planning appropriate levels of health care 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 due to[!] a) public/private insurance which means consumers rarely face market prices, b) agency relationships common in health between patients and doctors means it is unlikely observed consumption is based solely on patient preferences, and c) existence of interventions not yet in the market for which (by definition) market data do not exist. 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 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 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.

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 like 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 tradeoffs 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 (eg,[3,4]) and Information Integration Theory (eg,['l) in psychology; random utility theory based discrete choice models in economics (eg,[6]), discrete multivariate statistical models for contingency (crosstab) tables (eg,l'l) and the optimal design of statistical experiments (eg,[8]). DCEs were pioneered in marketing by Louviere and Woodworth (1983i[9]l, but quickly spread into other fields including applied economics, particularly transport (eg,[!OJ) and environmental economics (eg,[11]).

Since the first health application in the early 1990s[12], the number of studies has grown rapidly, see Ryan and Gerard (2003i[13]l and Ryan, Gerard and Amaya-Amaya (2008)[14] for reviews of the literature and method. Despite being popular, health applications have been criticised[". [16]; and while much of the critique by Bryan and Dolan (2004i"l was fair, Lancsar and Donaldson (2005i[17]l noted that their critique largely applied to early DCE health applications, and was not a critique or invalidation ofDCEs per se. This raises two important points: I) to some extent best practice in DCEs has been a moving target, and 2) it is unwise to apply DCEs without thoroughly understanding theory, methods 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 paper are to provide: a) an overview of basic DCE principles, b) guidance on key factors to consider in undertaking and assessing the quality ofDCE applications, including a detailed checklist, and c) 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.

To achieve these objectives the next section examines use of DCEs in economic evaluation in health to date. Section 3 presents their theoretical underpinnings. We discuss the process of designing, undertaking and interpreting DCEs in Section 4, where we highlight key factors in assessing the quality of these studies, summarised in a checklist. Section 5 traces out the research frontier and Section 6 concludes.

2. 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 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 and length of life, measured by Quality Adjusted Life Years (QALYsi[18,19], in CUA, to potentially capturing all forms of benefit (including health, non health and process benefits) using monetary valuation inCBA.

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 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 maximise more than QALYs[20] Also, it has been noted that QALYs measure health related utility only under specific restrictions on consumers' utility functions[21,22], 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 tradeoffs and predicting uptake to inform policy development and analysis; and 2) measuring outcomes for inclusion in economic evaluation. Initially applications focused on the first area, primarily eliciting patient preferences and tradeoffs for features describing products or programs in clinical settings (e.g.[23-25]), and also in broader contexts like GP's preferred remuneration packages[26], preferred health insurance packages (e.g.[27]), types of health service configurations (e.g.[28, 29]), and exploring time preference (e.g.[30]). 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 (e.g.[31-33]) and value health outcomes (e.g.[34, 35]).

Ahnost uniquely, DCEs potentially can provide inputs to both CBA and CUA, facilitating focus on the second of the two areas above. DCEs are increasingly used to elicit WTP for individual characteristics of goods/services and monetary measures of benefits as a whole for potential use in CBA[36, 37]. DCEs facilitate valuation of multiple options rather than evaluating a single intervention or treatment. The feasibility of eliciting utility weights from DCEs to be used to calculate QALYs is also being explored[34, 38, 39]. A possible advantage of using DCEs to elicit such weights is their grounding in utility theory. However, we are unaware ofDCE-derived outcome measures being used in CBA or CUA to date; we return to the potential expanded role for DCEs in economic evaluation in the penultimate section.

3. Theoretical basis

DCEs represent an integration of several theoretical areas. They are consistent with Lancaster's characteristics theory of demand: consumers have preferences for and derive utility from underlying attributes, rather than goods per se[40]. They also are consumer theory are preference-based and choice-based; the former assumes decision makers have a preference relation over a set of possible choices that satisfies certain axioms (completeness, transitivity, monotonicity, local non-satiation, convexity and

consistent with welfare and consumer theory[36,41]. The two main approaches to consumer theory are preference-based and choice-based; the former assumes 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 consistent with the weak axiom of revealed preference (see comparison in Lancsar and Louviere 2006[42]). The DCE approach to preference elicitation is akin to the choice-based approach to consumer theory as 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 4.

Choices made in DCEs are analysed using random utility theory (RUTi[6, 43]) which posits that utility for individual *i* conditional on choice *j* can be decomposed into an explainable component, V_{1}, and a non-explainable or random component ε,

$$U_{ij} = V_{ij} + \varepsilon_{ij} \qquad \textbf{(I)}$$

Economists view random components as due to unobservable or unobserved attributes, unobserved preference variation, specification error and/or measurement error[44]; psychologists view this component as being due to inherent variability within and between individuals[4]. The systematic component is a function of (at least) attributes of alternative and characteristics (covariates) of individual choosers, often modelled as

$$V_{ij} = X_{ij}'\beta + Z_{i}'\gamma \qquad (2)$$

where X_{ij} is the vector of attributes, usually including price and quality, of the *jth* good as viewed by the *ith* individual and Z, 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 I **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 $s_{,,}$ the probability that utility is maximised by choosing option I is given by

$$P(Y, = \mathbf{I}) = P(U'' > U,)$$
$$= P(V,, + \text{""} > V, + S,)$$
$$= P(V,, - V, > s, - \text{""}) \quad Vj * \blacksquare \qquad (3)$$

Where $Y;$ 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 as extreme value type I random variates, this results in a conditionallogit specification for the choice probabilities

$$P(Y_i = 1) = \frac{e^{\mu V_{i1}}}{\sum_{j=1}^{n} e^{\mu V_{ij}}} \quad j=i,\ldots,n \qquad (4)$$

Using equation (2), equation (4) can be rewritten as

$$P(Y_j = \mathrm{I}) = \frac{e''(x;, + z;,\mathbf{I}}{\mathbf{L}:=\mathbf{I}_{eJ.i.}(x,Jf3+Z,r}\mathbf{I}, \quad j = \mathbf{!},\ldots,n \qquad (5)$$

Equations (4) and (5) have an embedded scale parameter, $f.I,$ that is inversely proportional to the variance of the error distribution, a''; thus parameter estimates returned by estimation algorithms are $f3_{1aco,}$ not $f\}^{46}J$ $f.I$ cannot be identified in any one data source, so is usually set to one[^{11}J; 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[ll.^{45,47}J

4. Undertaking a DCE

DCEs involve three main inter-related components: I) an experimental design used to implement the choice survey and generate choice data; 2) discrete choice analysis to

estimate preferences from the choice data; and 3) use of the resulting model to derive welfare measures and conduct other policy analyses. We discuss each in tum and summarise these issues in a checklist provided in Table II.

4.1 Designing an experiment to generate choice data

Conceptualising 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[4s]

DCEs involve asking respondents to make discrete choices, in contrast with other SP methods like conjoint analysis ranking and rating tasks. Louviere and Lancsar (2008i[49]l 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 (eg, chiropractor, physiotherapy) or generic (eg, drug A, drug B). Labelled alternatives are specified in econometric analyses with alternative specific constants (ASCs, discussed below). 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[*so.*"l, 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 like lifestyle or other population health programs[31] From an evaluation perspective the comparator of interest is often a status quo treatment. If a status quo or opt out 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[52] 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 reported values used in model estimation (for example, see[32]).

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 summarised in sections I and 5 of Table II.

Define attributes and levels

Attributes can be quantitative (eg, waiting time) or qualitative (eg, provider of care) and are generally identified from literature, qualitative research like semi-structured interviews and/or focus groups with samples of relevant respondents, and experts (eg clinicians/policy makers[3]) DCEs 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 (2006)[42] 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 (eg "new horizon medications") by stretching level ranges. Indeed, a sufficiently wide range of levels should be used to

avoid respondents ignoring attributes due to 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, Skjoldborg and Gryd-Hansen (2003i[54]l found that changing the price vector changed parameter estimates and MRS (however, they note that changing the price vector compromised the experimental design, perhaps biasing results). In contrast, Hanley et a!(2005i"l 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 3). The payment vehicle (and duration) should be chosen to match the type of good and setting, which is well-known in contingent valuation[[56]] Special attention is required to properly describe risk attributes (e.g. risk of morbidity or mortality associated with different health states) as evidence suggests people may have difficulty interpreting probabilities[S?]

Types of attribute effects to be estimated also should 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 12 attributes[ss] In some settings achieving clinical relevance can required detailed attributes and levels thereby increasing complexity of the design. Researchers may inadvertently cause omitted variable bias by excluding key attributes, which should 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 summarised in sections 2, 3 and 5 of Table II.

Create 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 covariates like socio-demographics; so, 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 *LA* 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 and often is too large to be used in practice so a "fractional factorial" is typically used, which 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 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 factorials designs known as orthogonal main effects plans (OMEPs), implying a strictly additive IUF, typically have been used in health; which 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 2005[59] suggest this is not the case, although their simulation study used parameter values for non-linear terms and interactions that were so small that their IUFs were close to additive. More work is needed in this area. In the meantime larger fractional designs that allow estimation of (at least) all two-way interactions minimise 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 like 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 5 attributes, three with 4-levels and two with 2-levels, the full factorial produces 256 (4^3x2^2) combinations. This can be blocked into 16 versions of 16 choice sets with respondents randomly assigned to version. Fractional factorial designs also can be blocked into versions (eg,[32]). 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 insure 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 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-covariance 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

$$\left[\det(C)/\det(C_{opt})\right]^{1/p} \qquad (6)$$

where p is the number of parameters to be estimated in the model, C is defined above and Copt is the largest value of the C matrix. Street, Burgess and Louviere (2005i[60]) note 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 ofDCEs is entirely under the control of the researcher, so such identification problems should not occur.

Street and Burgess (2007i[6'l] developed theory to produce optimally or near optimally efficient designs for conditionallogit 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 2levels. 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, *LMA* designs are available, where *L* is number of levels, *A* number of attributes andM number of choice sets[39 ·62]

As noted earlier, it often is appropriate to include constant alternatives in choice sets like 'none of these' or status quo. Such options can reduce design efficiency, but typically this is outweighed by better congruency with consumer theory and grounding in reality. It is worth noting, however, that optimally efficient designs for generic choices also are optimal when a 'none of these' option is included[6ı]

In health, profiles are often obtained from software packages like SPEED, 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[6'l, 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[6ı]

Huber and Zwerina (1996i[63]l propose what they consider to be desirable design criteria: orthogonality (attribute levels appear in choice sets with equal frequency with each level of each other attribute), level balance (levels of each attribute appear equally often), minimum overlap (minimise the overlap of levels for each attribute in each choice) and utility balance (options in each choice set have similar probabilities of being chosen). Street and Burgess (2007)[6!] 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 (2005)[39] showed that utility balance can increase the variance of the error component, which, as highlighted in Section 3, can impact parameter estimates. Further, if all options in each set are approximately equal in utility, there would be no reliable statistical information for model estimation. Also some designs can 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 8[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[32]. Few studies have compared responses from individuals administered small versus large numbers of choice sets (e.g.[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 these factors increase unobserved variability at approximately a logarithmic rate[67]. So, 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 asthmatics to participate in all strenuous/sporting activity but does not allow participation in daily activities makes little sense. Possible solutions involve nesting attributes (eg, 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 tradeoffs 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, thereby requiring pilot testing. Key issues discussed above regarding the creation of an experimental design are summarised in section 4 of Table II.

Pilot tests

As with all primary data collection methods, iterative face to face pilot testing is needed to gnide development and testing ofDCE surveys. This includes 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 outlined above. We summarise these issues in section 6 of Table II.

Sample

Sampling requires considering the population to whom the results will be generalised, opportunity costs regarding how programs 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/program 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 tax payers 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** also is related to experimental design since the number of observations depends on

numbers of choice sets per respondent and numbers of respondents in the sample. A useful discussion of sampling for choice models (primarily for RP data) is provided in Ben-Akiva and Lerman (1985i[44]l; Louviere, Hensher and Swait (2000i[62]l 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 covariate effects, invariably requires larger sample sizes. See sections 7 and 8 of Table II.

Data collection

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

4.2 Discrete choice analysis

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-centring numerical attributes can be useful when specifying non-linear effects for numerical attributes like quadratic or cubic effects. Effects codes and mean-centering avoid correlations with the ASCs/intercepts, allows interpretation of the ASCs/intercepts as reflecting aggregate shares of choices and minimises collinearity in estimation

matrices used to estimate interactions. Several studies in health economics used effects codes (eg, [31, 32, 36]); the importance of using these codes is highlighted by Bech and Gyrd-Hansen (2005). 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 I shows an example of effects coding for a 4-level attribute. As can be seen in the table, this is very similar to dummy coding in that only L-1 levels (3 in this case) are coded, with the omitted L-th level on each effects coded variable coded -1 rather than 0. Coding is considered in section 10 of Table II.

Insert Table I

Forms of choice models

The form of the estimated IUF depends on 1) the experimental design and whether interaction effects are identified and/or alternatives are labelled, and 2) 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 [46]. Whether IIA holds is an empirical question, tests for which are outlined in Train (2002) [46].

Different choice models arise from different assumptions about distributions and properties of error components and about variance-covariance 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 (in health see for eg, [38]) and mixed logit (MIXL) (in health see for eg, [23, 58, 72]) to name two.

McFadden and Train (2000) [73] show that any random utility model can be approximated by a MIXL. MIXL has more flexible substitution patterns and accommodates 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 by including a respondent specific stochastic component: $\beta_i = \beta + \eta_i$, where β is the mean parameter vector for the population and η_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 (ie, mean and standard deviation). MIXL does not have a closed form solution, requiring simulated maximum likelihood estimation (or hierarchical hayes 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[45, 64, 65, 77, 78] 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 visualise implied functional forms. This allows recoding andre-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[79] For labelled DCEs with *J* alternatives, ASC for *J-1* alternatives should be included which represent the underlying preference for each alternative when attributes are set to zero. Naturally, specifications with interactions, socio-demographic variables and covariates 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 2) statistical considerations like 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 summarised in section II of Table II.

Validity

Validity ofDCEs is relatively well-established in the broader literature[[62]l, with comparisons to RP data in marketing, environmental and transportation economics[ll, [62- so] There have been relatively few tests of external validity in health, perhaps due to limited RP data, although Mark and Swait (2004i[8']l 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 (eg, completeness, monotonicity and transitivity[[82-84]l)_ Similarly, researchers also have 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 (2006)[[42] discuss several problems in testing "rationality", including the fact that apparent "irrationality" can be due to I) shortcomings in design and implementation ofDCEs; 2) respondent learning about their preferences or tasks; 3) "irrationality" tests not being conclusive; 4) using fractional factorials, which cannot identify unique decision rules_ They also provide 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 of issues to consider at each stage of undertaking/reviewing a DCE provided in Table II give more confidence in results_ Validity is considered in section 12 of Table IL

4,3 Interpretation, derivation of welfare measures & 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/program attributes_ For example, is test accuracy relatively more important to patients than time spent waiting for results when choosing diagnostic tests? Many studies measure the relative impact of attributes by comparing size and significance of estimated attribute parameters_ Unfortunately these parameters cannot be directly compared because attribute impacts and the positions of each attribute level on the underlying utility scale are confounded (ie, distances between utilities associated with attribute

levels need not be the same for each attribute[9]. To measure relative attribute impacts one needs to measure each on a common, comparable, scale. See section 13 of Table II. Lancsar, Louviere and Flynn (2007[9]) discuss five ways to compare relative attribute impact, some of which we discuss below.

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[31, 32, 79]. 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 also are 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 ofDCE results which is appropriate when market data are available.

Marginal rates of substitution (10RS)

DCEs allow estimation of tradeoffs that respondents make between attributes, or their MRS[24, 26, ss] 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:

$$MRS_{x_1 x_2} = \frac{\partial V / \partial X_1}{\partial V / \partial X_2} \qquad (7)$$

where Vis an IUF and X_1 and X_2 are attributes of the good/service. The numerator (denominator) is interpreted as the marginal utility of attribute I (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 (2) 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, Louviere and Flynn (2007).

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 was introduced to health economics to calculate welfare measures from DCEs by Lancsar (2002)[!] and Lancsar and Savage (2004i[36]l, and is due to Small and Rosen (1981)[86] The CV method can calculate measures of welfare gain, or WTP, for entire products/programs, and can measure the relative impacts of each attribute in a common monetary metric as willingness to pay or accept compensation for changes in a given attribute. For a conditionallogit model, both forms of welfare measures are calculated using the utility estimates and attribute levels in the following expression

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

where A is the marginal utility of income; V_1^0 and V_1^1 are the value of the IUF for each choice option j before and after the policy change, respectively; and J is the number of options in the choice set.

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[33] Equation (8) also can be used to calculate the CV using non monetary metrics; for example Baker eta!(2008)[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 (2004i[36]l note, 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 (2007i[87J provides a useful review and comparison of methods available to calculate confidence intervals. Issues to check regarding interpretation ofDCE results and welfare and policy analysis are included in sections 13 and 14 of Table II.

Insert Table II

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. Generalisability of results may be an issue in economic evaluation depending on how the DCE is designed and administered[". "l; a new DCE may be required for each research question, although due the flexibility ofDCEs several versions of a program or treatment can to be valued within a single study. We return to some of these issues in the next section.

5. 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 to develop DCEs, to (iteratively) apply 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 (2007)[[53]l, but qualitative methods remain underutilised.

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 how to optimise allocation of resulting choice sets into blocks. As DCEs become more complex, we need to better understand the relationships between design efficiency and respondent 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 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[44]

Analysis of DCEs in health has primarily focused on response means, but variances of outcome distributions and error components warrant attention[45,64,65] 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 3. Accounting for unobserved variability and preference heterogeneity is important, as are the relative merits of various ways to do this. Theory and methods have been developed to allow one to model the choices of single individuals, which eliminates a need to make distributional assumptions about preference heterogeneity[67]

DCEs can isolate and measure patient preferences, but decisions about health care 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 eta!(2006i [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 in decisions regarding water heaters. This approach could prove useful for similar

analysis in the health sector. Another extension includes modelling multi-stage choice processes which can include quantities chosen and changes in choices over time[45,65] Integration of more behavioural theory, and incorporating contributions from various fields like psychology should be beneficiaL

Generalisability ofDCE results is a key research need because it relates to factors like time, context and geography that often are constant in single data sources[93] Thus, we need to understand their effects, including 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" ofDCE responses, many tests focused on axioms not strictly required for rationality, 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[82,95]

A new type of choice experiment called "Best-Worst Scaling" (BWS) is garnering attention in health economics and more broadly, BWS' underlying theoretical properties were formally proven by Marley and Louviere (2005)[96] and Marley Louviere and Flynn (2008i[97] Thus far two types ofBWS 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[79,98,99]; and 2) 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[!OO] We expect such approaches to see increasing use in health economics,

While DCEs mainly have 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[36], suggesting they can be used in CBA; and Mcintosh (2006)[37] proposes 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 TTO, SG, VAS 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 potentially to be used directly in CUA (along the lines of a 'Super QALY'[lOll). 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 RCTs[32,102], and are likely to be included more routinely in RCTs as they become more mainstream in economic evaluation. There also may be a role for DCEs in the expanded remit of the National Institute for Health and Clinical Excellence (NICE) in the UK, which includes economic evaluations of population health initiatives since a key issue in such evaluations is participation/uptake, areas where DCEs can be useful. DCEs also may be useful for priority setting[103], including serving as a component of priority setting tools like program budgeting marginal analysis[l04J.

6. Conclusion

We reviewed and discussed the application and development ofDCEs in health economics more than 15 years after the first application. We also 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, 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, Lancsar and Louviere[''] 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 an accurate prediction, DCEs are a standard health economics research tool; they are not yet a standard health policy tool, but are starting to be used in that way.

Reference list

1. Lancsar E. Deriving welfare measures from stated preference discrete choice modelling experiments. CHERE Discussion Paper No 48: Centre for Health Economics Research and Evaluation, University of Technology, Sydney; 2002. Report No.: 48.
2. Drummond MF, Sculpher MJ, Torrance GW, eta!. Methods for the economic evaluation of health care programmes.. Third ed. Oxford: Oxford University Press; 2005.
3. Krantz DH, Tversky A. Conjoint measurement analysis of composition rules in psychology. Psychology Review. 1971;78:151-69.
4. Luce RD, Tukey JW. Simultaneous Conjoint Measurement: A New Type of Fundamental Measurement. Journal of Mathematical Psychology. 1964;I:1-27.
5. Anderson NH. Foundations oflnforrnation Integration Theory. New York: Academic Press; 1981.
6. McFadden D. Conditionallogit analysis of qualitative choice behavior. In: Zarembka P, editor. Frontiers of econometrics. New York: Academic Press; 1974. p. 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, NY: Wiley; 1978.
9. Louviere J, Woodworth G. Design and analysis of simulated consumer choice or allocation experiments: an approach based on aggregated data. Journal of Marketing Research. 1983;20:350-67.
10. Hensher DA, Louviere JJ. On the design and analysis of simulated choice or allocation experiments in travel choice modelling. Transportation Research. 1983;890:11-7.
11. Adamowicz W, Louviere J, Williams M. Combining revealed and stated preference methods for valuing environmental amenities. Journal of Environmental Management and Economics. 1994;26(3):271-92.
12. Propper C. Contingent valuation of time spent on NHS waiting lists. Economic Journal. 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 Economics and 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. BryanS, Dolan P. Discrete choice experiments in health economics. For better or for worse?[see comment]. European Journal of Health Economics. 2004;5(3):199-202.
16. Wainright DM. More 'con' than 'joint': problems with the application of conjoint analysis to participatory healthcare decision making. Critical 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]. European Journal of Health Economics. 2005;6(4):314-6.
18. Torrance GW. Measurement of health state utilities for economic appraisal. Journal of Health Economics. 1986 5:1-30.
19. Williams A. Economics of coronary artery bypass grafting. British Medical Journal. 1985 291:326-9.

20. Nord E, Pinto JL, Richardson J, eta!. Incorporating societal concerns for fairness in numerical valuations of health programmes.[see comment][erratum appears in Health Econ 1999 Sep;8(6):559]. Health Economics. 1999;8(1):25-39.
21. Bleichrodt H. QALYs and HYEs (healthy year equivalents): under what conditions are they equivalent? Journal of Health Economics. 1995;14(1):17-37.
22. Pliskin J, Shepard D, Weinstein, W. Utility functions for life years and health status
Operations research. 1980;28:206-24.
23. Johnson FR, BanzhafMR, Desvousges WH. Willingness to pay for improved respiratory and cardiovascular health: a multiple-format, stated-preference approach. Health Economics. 2000;9(4):295-317.
24. Ryan M. Using conjoint analysis to take account of patient preferences and go beyond health outcomes: an application to in vitro fertilisation. Social Science and Medicine. 1999;48(4):535-46.
25. Ryan M, Hughes J. Using conjoint analysis to assess women's preferences for miscarriage management. Health Economics. 1997;6:261-73.
26. Scott A. Eliciting GPs' preferences for pecuniary and non-pecuniary job characteristics. Journal of Health Economics. 2001;20:329-47.
27. Chakraborty G, Ettensen R, Gaeth G. How consumers choose health insurance. Journal of Health Care Market. 1994;14:21-33.
28. Jan S, Mooney G, Ryan M, eta!. 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.
29. Morgan A, Shackley P, Pickin M, et a!. Quantifying patient preferences for out-of-hours primary care. Journal of Health Services Research and Policy. 2000;5:214-18.
30. van der Pol M, Cairns J. Estimating time preference for health using discrete choice experiments. Social Science and Medicine. 2001;52:1459-70.
31. Hall J, Kenny P, King M, et a!. Using stated preference discrete choice modelling to evaluate the introduction of varicella vaccination. Health Economics. 2002;11:457-65.
32. King MT, Hall J, Lancsar E, eta!. Patient preferences for managing asthma: results from a discrete choice experiment. Health Economics. 2007;16(7):703-17.
33. Lancsar EJ, Hall JP, King M, eta!. Using discrete choice experiments to investigate subject preferences for preventive asthma medication. Respirology. 2007;12(1):127-36.
34. Hakim Z, Pathak DS. Modelling the EuroQol data: a comparison of discrete choice conjoint and conditional preference modelling. Health Economics. 1999;8(2):103-16.
35. Sculpher M, BryanS, Fry P, eta!. Patients' preferences for the management of non-metastatic prostate cancer: discrete choice experiment BMJ. 2004;328:382-4.
36. Lancsar E, Savage E. Deriving welfare measures from discrete choice experiments: inconsistency between current methods and random utility and welfare theory.[see comment]. Health Economics. 2004;13(9):901-7.
37. Mcintosh E. Using discrete choice experiments within a cost-benefit analysis framework: some considerations. Pharrnacoeconomics. 2006;24(9):855-68.
38. Ryan M, Netten A, Skatun D, eta!. Using discrete choice experiments to estimate a preference-based measure of outcome--an application to social care for older people. Journal of Health Economics. 2006;25(5):927-44.

39. Viney R, Savage E, Louviere J. Empirical investigation of experimental design properties of discrete choice experiments in health care. Health Economics. 2005;14(4):349-62.
40. Lancaster K. A new approach to consumer theory. Journal of Political Economy. 1966;74:132-57.
41. Hanley N, Mourato S, Wright RE. Choice modelling approaches: A superior alternative for envirornnental valuation? Journal of Economic Surveys. 2001;15:435-62
42. Lancsar E, Louviere J. Deleting 'irrational' responses from discrete choice experiments: a case of investigating or imposing preferences? Health Economics. 2006;15(8):797-811.
43. Thurstone L. A law of comparative judgement. Psychological Review. 1927;34:273-86.
44. Ben - Akiva M, Lerman SJ. Discrete Choice Analysis: Theory and Applications to Travel Demand. Cambridge: The MIT Press; 1985.
45. Adamowicz W, Bunch D, Cameron T, eta!. Behvioural frontiers in choice modelling. Marketing Letters. In press.
46. Train KE. Discrete Choice Methods with Simulation. Cambridge: Cambridge University Press; 2003.
47. Swait J, Louviere J. The Role of the Scale Parameter in the Estimation and Comparison of Multinomial Logit Models. Journal of Marketing Research. 1993;30:305-14.
48. Carson RT, Groves T, Machina MJ. Incentive and Informational Properties of Preference Questions. San Diego: University of California, San Diego; 2000.
49. Louviere J, Lancsar E. Distinguishing between conjoint analysis and discrete choice experiments with implications for stated preference and welfare elicitation. Sydney: CenSoC, University of Technology, Sydney; 2008.
50. Ryan M, Skatun D. Modelling non-demanders in choice experiments. Health Economics. 2004;13(4):397-402.
51. Viney R, Lancsar E, Louviere J. Discrete Choice Experiments to Measure Consumer Preferences for Health and Healthcare. Expert Review of Pharmacoeconomics and Outcomes Research. 2002 August;2(4):319-26.
52. Kahneman D, Tversky A. Prospect Theory: An Analysis of Decision under Risk. Econometrica. 1979;47:263-91.
53. Coast J, Horrocks S. Developing attributes and levels for discrete choice experiments using qualitative methods. Journal of Health Services & Research Policy. 2007;12(1):25-30.
54. Slothuus Skjoldborg U, Gyrd-Hansen D. Conjoint analysis. The cost variable: an Achilles' heel? Health Economics. 2003;12(6):479-91.
55. Hanley N, Adamowicz W, Wright RE. Price vector effects in choice experiments: an empirical test. Resource and Energy Economics. 2005;27:227-34.
56. Smith R. Construction of the contingent valuation market in health care: a critical assessment. Health Economics. 2003;12:609-28.
57. Peters E, Vastfjall D, Slovic P, eta!. Numeracy and Decision Making. Psychological Science. 2006;17(5):407-13.
58. Hall J, Fiebig DG, King MT, eta!. What influences participation in genetic carrier testing? Results from a discrete choice experiment. Journal of Health Economics. 2006;25(3):520-37.
59. Lusk JL, Norwood FB. American Journal of Agricultural Economics. 87. 2005;3(771-785).

60. Street DA, Burgess L, Louviere JJ. Quick and Easy Choice Sets: Constructing Optimal and Nearly Optimal Stated Choice Experiments International Journal of Research in Marketing. 2005;22:459- 70.
61. Street DA, Burgess L. The Construction of Optimal Stated Choice Experiments: Theory and Methods. Hoboken, New Jersey: Wiley; 2007.
62. Louviere JJ, Hensher, D.A, Swait, J.D. Stated Choice Methods Analysis and Application. Cambridge: Cambridge University Press; 2000.
63. Huber J, Zwerina K. The importance ofutlity balance in efficient choice designs. Journal of Marketing Research. 1996;33:307-17.
64. Louviere J, Engle T. Confound it! That pesky little scale constant messes up our convenient assumptions. Sawtooth Software Conference 2006; Dequm, USA; 2006. p. 211-28.
65. Louviere JJ, Meyer RJ. Formal Choice Models oflnforrnal choices: What Choice Modeling Research Can (and Can't) Learn from Behavioural Theory. Review of Marketing Research. In press.
66. DeShazo JR, Fenno G. Designing choice sets for stated preference methods: the effects of complexity on choice consistency. Journal of Environmental Economics and Management 2002;44:123-43.
67. Louviere J, Islam T, Wasi N, eta!. Designing discrete choice experiments: Do optimal designs come at a price? Journal of Consumer Research. In press.
68. San Miguel F, Ryan M, Amaya-Amaya M. Irrational stated preferences: a quantitative and qualitative investigation. Health Economics. 2004;14:307-22.
69. Shackley P, Donaldson C. Willingness to pay for publicly-financed health care: how should we use the numbers? . Applied Economics. 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 Economics. 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. Journal of 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. Journal of Consumer Research. 2001;28(1):135-48.
75. Islam T, Louviere JJ, Burke PF. Modeling the Effects oflncluding!Excluding Attributes in Choice Experiment on Systematic and Random Components. International Journal of Research in Marketing. 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. Organizational Behavior & Human Decision Processes. 2001;86(2):141-67.
77. Fiebig DG, Keane M, Louviere J, eta!. The generalised multinomiallogit model. Marketing Science. Conditionally accepted.
78. Magidson J, Verrnunt J. Removing the Scale Factor Confound in Multinomial Logit Choice Models to Obtain Better Estimates of Preference. Sawtooth Software Conference, 2008; 2007 November 2007; 2007.

79. Lancsar E, Louviere J, Flynn T. Several methods to investigate relative attribute impact in stated preference experiments. Social Science & Medicine. 2007;64(8):1738-53.
80. Adamowicz W, Swait J, Boxall P, eta!. Perceptions versus objective measures of environmental quality in combined revealed and stated preference models of environmental valuation. Journal of Environmental and Management. 1997;32:65-84.
81. Mark T, Swait J. Using stated preference and revealed preference modelling to evaluate prescribing decisions. Health Economics. 2004;13(6):563-73.
82. Lloyd AJ. Threats to the estimation of benefit: are preference elicitation methods accurate? Health Economics. 2003;12:393-402.
83. Mcintosh E, Ryan M. Using discrete choice experiments to derive welfare estimates for the provision of elective surgery: Implications of discontinuous preferences. Journal of Economic Psychology. 2002 2002/6;23(3):367-82.
84. Ryan M, San Miguel F. Revisiting the axiom of completeness in health care. Health Economics. 2003;12(4):295-307.
85. Gyrd-Hansen D, Sogaard J. Analysing public preferences for cancer screening programmes. Health Economics. 2001;10(7):617-34.
86. Baker R, Donaldson C, Lancsar E, eta!. Deriving QALY weights through discrete choice experiments: challenges and preliminary results. Health Economics Study Group Meeting. Norwich; 2008.
87. Risa Hole A. A Comparison of Approaches to Estimating Confidence Intervals for Willingness to Pay Measures. Health Economics. 2007;16:827-40.
88. Maddala T, Phillips KA, Reed Johnson F. An experiment on simplifying conjoint analysis designs for measuring preferences. Health Economics. 2003;12(12):1035-47.
89. Severin V. Comparing statistical efficiency and respondent efficiency in choice experiments. Sydney: University of Sydney; 2000.
90. Bryan S, Gill P, Greenfield S, eta!. The myth of agency and patient choice in health care? The case of drug treatments to prevent coronary disease. Social Science & Medicine. 2006;63(10):2698-701.
91. VickS, Scott A. Agency in health care. Examining patients' preferences for attributes of the doctor-patient relationship. Journal of Health Economics. 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 Economics. 2006;31:639-391.
93. Louviere J, Street D, Carson R, eta!. Dissecting the Random Component of Utility. Marketing Letters. 2002 2002/8;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. Journal of Health Economics. 2005;24:591-611.
95. Swait J. A non-compensatory choice model incorporating attribute cutoffs. Transportation Research Part B: Methodological. 2001 2001/11;35(10):903-28.
96. Marley A, Louviere J. Some probabilistic models of best, worst, and best-worst choices. Journal of Mathematical Psychology. 2005;49(6):464- 80
97. Marley AAJ, Louviere JJ, Flynn T. Probabilistic Models of Set-Dependent and Attribute-Level Best-Worst Choice. Journal of Mathematical Psychology. In press.
98. Flynn T, Louviere J, Peters T, eta!. Best-Worst Scaling: What it can do for health care and how to do it. Journal of Health Economics. 2007;26 171-89.

99. Mcintosh E, Louviere J. Separating weight and scale value: an exploration of best-attribute scaling in health economics. Health Economics Study Group Meeting, Brunei University; 2002 January 2002; 2002.
100. 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. Newcastle upon Tyne, UK; 2005.
101. Anand P. QALYs and the integration of claims in health-care rationing. Health Care Analysis. 1999;7:239-53.
102. Longo MF, Cohen DR, Hood K, eta!. Involving patients in primary care consultations: assessing preferences using discrete choice experiments. British Journal of General Practice. 2006;56:35-42.
103. Ratcliffe J. Public preferences for the allocation of donor liver grafts for transplantation. Health Economics. 2000;9(2):137-48.
104. Ruta D, Mitton C, Bate A, eta!. Programme budgeting and marginal analysis: bridging the divide between doctors and managers. British Medical Journal. 2005;330:!501-3.

Tables

Table 1: Effects coding for a four level attribute

Levels	Fxl	Fx2	Fx3
0	I	0	0
I	0	I	0
2	0	0	I
3	-I	-I	-I