VALUE IN HEALTH 16 (2013) 3–13

Available online at www.sciencedirect.com

SciVerse ScienceDirect

journal homepage: www.elsevier.com/locate/jval

ISPOR TASK FORCE REPORT

Constructing Experimental Designs for Discrete-Choice Experiments: Report of the ISPOR Conjoint Analysis Experimental Design Good Research Practices Task Force

F. Reed Johnson, PhD[1,*], *Emily Lancsar, PhD*[2], *Deborah Marshall, PhD*[3], *Vikram Kilambi, BA/BS*[1], *Axel Mühlbacher, PhD*[4,5], *Dean A. Regier, PhD*[6], *Brian W. Bresnahan, PhD*[7], *Barbara Kanninen, PhD*[8], *John F.P. Bridges, PhD*[9]

[1]*Health Preference Assessment Group, RTI Health Solutions, Research Triangle Park, NC, USA;* [2]*Centre for Health Economics, Faculty of Business and Economics, Monash University, Melbourne, Victoria, Australia;* [3]*Faculty of Medicine, Department of Community Health Sciences, University of Calgary, Calgary, Alberta, Canada;* [4]*Duke University, Durham NC, USA;* [5]*Hochschule Neubrandenburg, Neubrandenburg, Germany;* [6]*Canadian Centre for Applied Research in Cancer Control, British Columbia Cancer Agency, Vancouver, Canada;* [7]*Department of Radiology, University of Washington, Seattle, WA, USA;* [8]*BK Econometrics, LLC, Arlington, VA, USA;* [9]*Department of Health Policy and Management, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, USA*

ABSTRACT

Stated-preference methods are a class of evaluation techniques for studying the preferences of patients and other stakeholders. While these methods span a variety of techniques, conjoint-analysis methods—and particularly discrete-choice experiments (DCEs)—have become the most frequently applied approach in health care in recent years. Experimental design is an important stage in the development of such methods, but establishing a consensus on standards is hampered by lack of understanding of available techniques and software. This report builds on the previous ISPOR Conjoint Analysis Task Force Report: Conjoint Analysis Applications in Health—A Checklist: A Report of the ISPOR Good Research Practices for Conjoint Analysis Task Force. This report aims to assist researchers specifically in evaluating alternative approaches to experimental design, a difficult and important element of successful DCEs. While this report does not endorse any specific approach, it does provide a guide for choosing an approach that is appropriate for a particular study. In particular, it provides an overview of the role of experimental designs for the successful implementation of the DCE approach in health care studies, and it provides researchers with an introduction to constructing experimental designs on the basis of study objectives and the statistical model researchers have selected for the study. The report outlines the theoretical requirements for designs that identify choice-model preference parameters and summarizes and compares a number of available approaches for constructing experimental designs. The task-force leadership group met via bimonthly teleconferences and in person at ISPOR meetings in the United States and Europe. An international group of experimental-design experts was consulted during this process to discuss existing approaches for experimental design and to review the task force's draft reports. In addition, ISPOR members contributed to developing a consensus report by submitting written comments during the review process and oral comments during two forum presentations at the ISPOR 16th and 17th Annual International Meetings held in Baltimore (2011) and Washington, DC (2012).

Keywords: conjoint analysis, discrete-choice experiment, experimental design, stated preferences.

* *Address correspondence to:* F. Reed Johnson, Health Preference Assessment Group, RTI Health Solutions, 200 Park Offices Drive, Research Triangle Park, NC 27709, USA.

E-mail: frjohnson@rti.org.

1098-3015/$36.00 – see front matter Copyright © 2013, International Society for Pharmacoeconomics and Outcomes Research (ISPOR). Published by Elsevier Inc.

http://dx.doi.org/10.1016/j.jval.2012.08.2223

Background to the Task Force

The ISPOR Conjoint Analysis Experimental Design Task Force is the second ISPOR Conjoint Analysis Task Force. It builds on a previous task force report, Conjoint Analysis Applications in Health—A Checklist: A Report of the ISPOR Good Research Practices for Conjoint Analysis Task Force [1]. The previous report developed a 10-point checklist for conjoint analysis in the following areas: 1) the research question, 2) the attributes and levels, 3) the format of the question, 4) the experimental design, 5) the preference elicitation, 6) the design of the instrument, 7) the data-collection plan, 8) the statistical analysis, 9) the results and conclusions, and 10) the study's presentation [1,2].

The task force determined that several items, including experimental design, deserved more detailed attention. A proposal was developed to focus on experimental design to assist researchers in evaluating alternative approaches to this difficult and important element of a successful conjoint-analysis study. The ISPOR Conjoint Analysis Experimental Design Task Force proposal was submitted to the ISPOR Health Science Policy Council in October 2010. The council recommended the proposal to the ISPOR Board of Directors, and it was subsequently approved in November 2010.

Researchers experienced in experimental design, stated preferences, and discrete-choice experiments working in academia and research organizations in Germany, Australia, Canada, and the United States were invited to join the task force's leadership group. The leadership group met via bimonthly teleconference to identify and discuss current experimental-design techniques, develop the topics and outline, and prepare draft manuscripts. An international group of experimental-design experts was consulted during this process to discuss existing approaches for experimental design and to review the task force's draft reports.

The task force met in person at ISPOR International Meetings and European Congresses as well as held a task force face-to-face consensus meeting in March 2012 to come to agreement on several outstanding issues. In addition, ISPOR members contributed to developing a consensus report. The ISPOR Conjoint Analysis Review Group submitted written comments during the review process and oral comments during two forum presentations at the ISPOR 16th and 17th Annual International Meetings held in Baltimore (2011) and Washington, DC (2012).

Introduction

Stated-preference methods in the form of discrete-choice experiments (DCEs) are increasingly used in outcomes research as a means to identify and evaluate the relative importance of aspects of decision making related to health outcomes and health care services. Stated-preference methods are a class of evaluation techniques used for studying the preferences of patients and other stakeholders [3]. While these methods span a variety of techniques, conjoint-analysis methods—and particularly DCEs—have become the most frequently applied approach in health care in recent years [4,5].

This report builds on the previous ISPOR Conjoint Analysis Task Force Report: Conjoint Analysis Applications in Health—A Checklist: A Report of the ISPOR Good Research Practices for Conjoint Analysis Task Force [1]. This earlier report provides the steps to take for the development, analysis, and publication of conjoint analyses. The authors determined that several steps, including experimental design, deserved more detailed attention.

Experimental design refers to the process of generating specific combinations of attributes and levels that respondents evaluate in choice questions. The previous task force report indicated that good research practice requires researchers to evaluate alternative experimental-design approaches and justify the particular approach chosen [1]. Unfortunately, many researchers do not provide adequate documentation of the experimental design used in their studies. Poor support for the selected design strategy could indicate a lack of awareness of the applicability of alternative approaches for a given study. There have been significant advances, as well as significant confusion, in experimental-design methods in recent years. This report provides researchers with a more detailed introduction to constructing experimental designs on the basis of study objectives and the statistical model researchers have selected for the study.

The Conjoint Analysis Experimental Design Task Force report differs from the earlier task force report by limiting attention to one aspect of conjoint analysis—experimental design— and focuses specifically on one preference-elicitation method, DCEs. Furthermore, while the earlier report was directed at researchers with limited experience with conjoint-analysis methods, the topic of experimental design requires familiarity with these methods, as well as an awareness of some of the basic principles related to experimental design. For this background information, readers are directed to several systematic reviews of conjoint-analysis applications in health care [5–13] and several methodological reviews [6,14–18]. In this report, we provide some background on

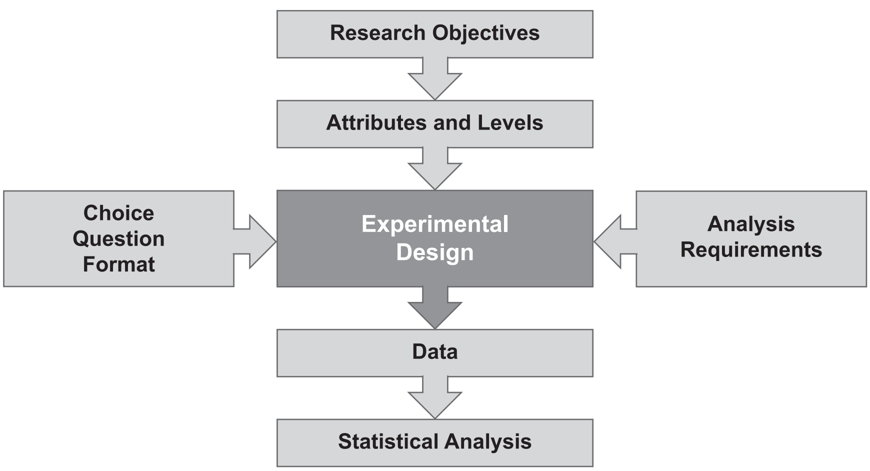

Fig. 1 – Key stages for developing a discrete-choice experiment.

the DCE approach and experimental design, but we advise readers who are interested in obtaining a deeper understanding of these concepts to consult the primary references in the field [8,19–23].

The Role of Experimental Design

Figure 1 illustrates where the experimental design fits into the key stages of developing a DCE. At each stage, researchers are required to select among several research approaches. *Research Objectives* refer to the construct, commodity, health condition, health care program, or other object of choice for which preferences will be quantified. *Attributes and Levels* are the individual features that comprise the research object, among which the survey will elicit trade-offs. Attributes may include such features as effectiveness, safety, or mode of administration of a pharmaceutical, biological treatment, or medical device; attribute levels describe the possible values, outcomes, interventions, or technologies associated with each attribute. For example, a service attribute could include levels of service quality or waiting time to receive care from a health care professional.

The *Choice Question Format* describes how a series of sets of alternatives from among all the possible profiles of attribute-level combinations will be presented to respondents. *Analysis Requirements* encompass information about the intended choice-model specifications. The *Attributes and Levels, Choice Question Format,* and *Analysis Requirements* all form the basis for the *Experimental Design*—which is subsequently used to construct the choice questions that are shown to respondents.

Data from the choice questions are then analyzed to predict choice and produce estimated preference weights, or choice-model parameters, that are consistent with the observed pattern of choices by respondents (*Statistical Analysis*). The resulting estimates are then used to evaluate treatment or policy options related to the research object. This report from the ISPOR Conjoint Analysis Experimental Design Task Force focuses on *Experimental Design*, represented by the black box in Figure 1. The previous Task Force report on conjoint-analysis methods discussed strategies for determining research objectives and specifying the attributes and levels [1].

In a DCE study, researchers use an experimental design to map attributes and levels into sets of alternatives to which respondents indicate their choices. As indicated in Figure 1, the experimental design comes after researchers have determined whose preferences (patients, caregivers, or providers) are being assessed, what health care features are of interest, and what types of models will be used. Experimental designs thus first require the researcher to determine the objectives of the study and to select the component attributes that are believed to characterize the health care object of interest. This, in turn, requires the following considerations:

- An explicit specification of the features (attributes) of a health care intervention to be tested for a particular stakeholder of interest;
- The specific type of value and range of values (levels) over which these features will be tested (e.g., duration of 2–4 weeks, 5%–10% chance of efficacy);
- The way in which observations, choices, or judgments made from among the alternatives will be presented and recorded; and
- A strategy for how the observed data will be modeled as a function of the attributes, levels, and other factors.

The experimental-design step consists of defining a systematic plan that determines the content of the choice questions to generate the variation in the attribute levels required to elicit a choice response. Efficient experimental designs maximize the precision of estimated choice-model parameters for a given number of choice questions.

While this report restricts itself to DCEs, the methods and procedures described here are applicable to other domains of stated-preference research. Approaches such as multiprofile best-worst scaling use these approaches to construct experimental designs that maximize statistical efficiency for choice questions in which respondents are asked to choose the best and worst outcomes, technologies, or interventions from a list [24–27]. Some principles relevant to DCE experimental designs, however, may not apply to other types of best-worst scaling formats. Moreover, researchers who are interested in combining both stated- and revealed-preference data may find general knowledge of experimental design useful when creating designs for this purpose [28].

Experimental-Design Concepts

Model Identification

Much attention has been paid in the recent literature to statistical efficiency in constructing experimental designs for choice experiments [11,15,16,29,30]. The first and most important consideration for a researcher, however, is identification. *Identification* refers to the ability to obtain unbiased parameter estimates from the data for every parameter in the model. Generating a design that allows for statistical identification of every parameter of interest requires researchers to specify a choice model (with every parameter coded) and to ensure that sufficient degrees of freedom are available for estimation.

Street and Burgess [16] noted that a number of designs used in studies found in the health care literature had identification problems; in particular, some studies had one or more effects that were perfectly confounded with other effects, meaning that the effects could not be independently identified and could produce biased estimates. Louviere and Lancsar [11] advised, "Given our current knowledge about the consequences of violating maintained assumptions associated with designs … we recommend that one first focus on identification, and then on efficiency, because one may be able to improve efficiency by increasing sample size, but identification cannot be changed once a design is constructed."

In general, the model specification, the number of attributes, and the functional form of attributes determine the numbers and types of parameters to be estimated. The review by Marshall et al. [6] estimated that 70% of the studies used three to seven attributes, with most studies having six attributes, or four attribute levels. Health outcomes, interventions, or technologies sometimes can be described by a continuous scale, such as blood pressure or time spent in a waiting room, but often can be described only by discrete, categorical end points, such as tumor stage, mode of administration, or qualitative severity indicators (such as "mild," "moderate," or "severe"). Categorical variables increase the number of parameters that must be estimated for each attribute.

In the case of continuous variables, researchers must specify levels to assign to the design, but the variable can be assumed to have a linear effect in the model—one parameter (for constant marginal utility) applied to all levels of the variable. Under the assumption of linear effects, designs based on categorical variables actually "overidentify" the model—that is, they use more degrees of freedom than necessary to model linear effects. Because such designs allow estimating a separate parameter for every level of the attribute, a smaller design with fewer choice questions could identify the intended statistical model. If the correct functional form is uncertain, however, an advantage of categorical variables is that they allow researchers to test and examine a variety of continuous specifications, including linear or nonlinear models, after data have been collected.

To identify particular effects of interest, the experimental design must sufficiently vary the relevant attribute levels within and across choice questions and, in the case of higher-order effects, include sufficient numbers of attribute-level combinations. As a simple example, consider an experiment in which researchers are interested in understanding how effectiveness (no pain vs. mild pain) and serious side effects (risk of myocardial infarction vs. risk of infection requiring hospitalization) affect treatment preferences. Suppose researchers ask respondents to choose between 1) a treatment that has no pain and a risk of myocardial infarction and 2) a treatment that has mild pain and a risk of infection. Suppose also that respondents tend to choose the treatment that has mild pain and a risk of infection. Did respondents select this option because of the acceptable effectiveness of the treatment or to avoid a particular side effect?

Researchers cannot distinguish between the independent effects of the effectiveness and side-effect attributes without observing choices for additional treatment combinations. Specifically, in this case, researchers need to add to the choice sets a choice between 1) a treatment with mild pain and a risk of myocardial infarction and 2) a treatment with no pain and a risk of infection. Adding these alternatives will allow researchers to observe how respondents react to varying each attribute independently. Certain classes of experimental designs, including orthogonal designs (discussed in the following sections), have the desirable property of independent variation by requiring that correlations among attributes all be zero.

In many cases, researchers are interested in estimating interaction effects. Estimating all interactions (two-way, three-way, and higher-order interactions) requires large, full-choice designs that include the complete set of combinations of all the attribute levels. These designs may include implausible combinations (prior to applying any restrictions) and generally are quite large, requiring often impractically large sample sizes and/or numbers of choice questions posed to each respondent.

For example, a two-alternative design using four attributes, each with three levels, yields 81 (3^4) possible profiles and has 3240 possible combinations of two-alternative choice questions [$3^4 \times (3^4 - 1)/2$]. Note that the number of feasible choice questions is less than the full factorial of all possible combinations of attribute levels. The full factorial includes pairing attribute levels with themselves, for example. If researchers are interested in only main effects or in a subset of possible interactions, then these models can be estimated by using a much smaller fraction of the full-choice design.

Whether to include or not to include interaction terms generally requires consideration of theory, intuition, and feasibility in terms of sample size and survey-design parameters. Not including interaction terms imposes the assumption, a priori, that such interactions are not statistically significantly different from zero or, if they are significant, that they are independent of the remaining attribute effects. However, this assumption may not be true, in which case the interaction effects are confounded with the main effects and the resulting estimates are biased. In the past, researchers often used main-effects designs for simplicity and feasibility, and so any resulting confounding and bias were accepted as an unavoidable consequence of this choice. Newer methods and software, however, can easily construct designs that accommodate more complex model specifications.

Statistical Efficiency versus Response Efficiency

In most studies, researchers do not necessarily have one, specific estimate of interest; rather, they would like to obtain a set of parameter estimates that jointly are as precise as possible. *Statistical efficiency* refers to minimizing the confidence intervals around parameter estimates in a choice model for a given sample size. Perfectly efficient designs are balanced, meaning that each level appears equally often within an attribute, and orthogonal, meaning that each pair of levels appears equally often across all pairs of attributes within the design.

Unlike revealed-preference methods, stated-preference methods allow researchers to control the stimuli that generate the data. As a result, some experts insist that experimental designs satisfy a very high standard for statistical efficiency. While statistical efficiency is the primary focus of most of the experimental-design literature, the overall precision of the resulting parameter estimates depends on both statistical efficiency and response efficiency. *Response efficiency* refers to measurement error resulting from respondents' inattention to the choice questions or other unobserved, contextual influences. Various cognitive effects that result in poor-quality responses to the experimental stimuli can cause measurement error. Some possible sources of measurement error include the following:

- Simplifying decision heuristics used by respondents that are inconsistent with utility maximization or the presumed choice model;
- Respondent fatigue resulting from evaluating a large number of choice questions;
- Confusion or misunderstanding or unobserved, heterogeneous interpretation by respondents, resulting from poorly constructed attribute and attribute-level definitions; and
- Respondent inattention resulting from the hypothetical context of the study

While measurement error cannot always be controlled for, it can be reduced by adherence to best survey-research practices; there may be study-design trade-offs between maximizing statistical efficiency and maximizing response efficiency. Statistical efficiency is improved by asking a large number of difficult trade-off questions, while response efficiency is improved by asking a smaller number of easier trade-off questions. Maximizing the overall precision of the estimates requires balancing these two sources of potential error [31].

Statistical efficiency and the ability to ask a large number of trade-off questions depend on the intended sample size. Confidence intervals shrink as a function of the inverse of the square root of the sample size. Sample sizes in the range of 1000 to 2000 respondents thus will produce small confidence intervals, even if the experimental design is not particularly efficient. However, many health studies have research-resource constraints or involve fairly rare conditions that limit sample sizes to 100 to 300 respondents [6]. In those circumstances, efficient experimental designs are critical to the success of the study.

Figure 2 is a plot of the effect of simulated sample sizes on estimate precision for three DCE studies [32]. Researchers sampled with replacement from each data set to simulate sample sizes ranging from 25 to 1000. A conditional-logit model was estimated for each of the 10,000 draws for each sample size, and a summary measure of estimation precision was calculated. The vertical axis is the mean of these calculations.

For all studies, precision increases rapidly at sample sizes less than 150 and then flattens out at around 300 observations. Differences in precision among studies converge for large sample sizes. While the shape of the plots in Figure 2 indicates that precision varies as expected with the inverse of the square root of sample size, variations in the positions of the study plots suggest that the effect of measurement error varies across studies. For example, a precision of 0.5 was obtained at a sample size of 250 for the cancer-screening study. The same level of precision required 600 observations for the platelet-disorder study. Thus,

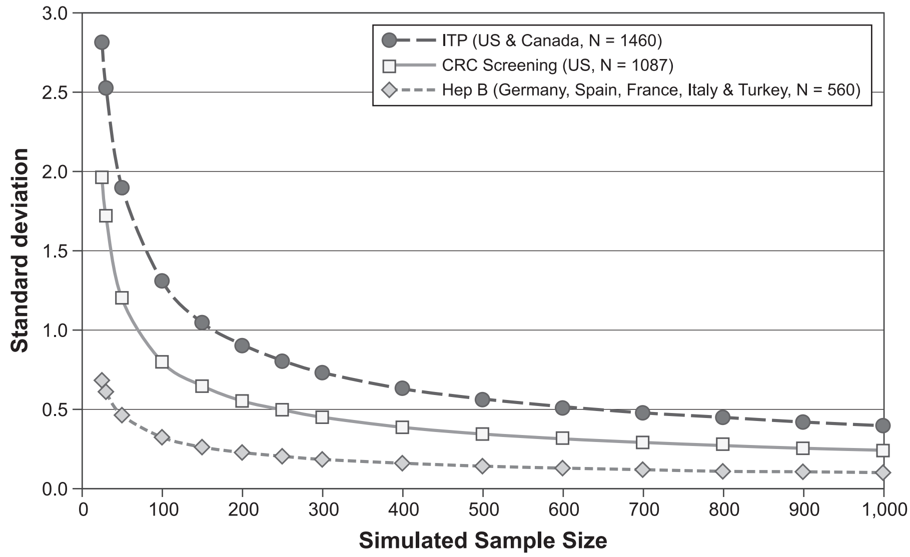

Fig. 2 – Effect of sample size on estimate precision.

for any given level of precision, measurement error can have a significant effect on the required sample size.

Experimental-Design Challenges in Health Applications

All applications of DCE methods require developing experimental designs. Health care studies, however, involve a number of technical considerations. In particular, many health applications involve such concerns as implausible attribute-level combinations; interaction effects among health outcomes, technologies, or interventions; cognitive limitations of some respondent groups; the role of labeled and constant alternatives; and blocking.

Potential implausible combinations

Because choice data are collected by using health profiles based on hypothetical alternatives, some possible attribute-level combinations could be implausible or illogical. An example of an implausible combination, or one that is inconsistent with logical expectation, would be a design with two attributes, *activities of daily living* (no restrictions vs. some restrictions) and *symptoms* (mild vs. moderate vs. severe). A DCE question that asks a respondent to evaluate a treatment alternative that combines no restrictions with severe symptoms would result in an implausible scenario or outcome. Respondents will have difficulty evaluating such illogical combinations, which could increase the potential for hypothetical bias; unobserved, heterogeneous interpretations by respondents; or lower response efficiency. Some design approaches allow researchers to specify combinations that should not appear in the design, while other approaches do not.

Interaction effects

In health application–specific research questions, the associated list of relevant attributes may include scenarios where interactions are likely among different attributes. In particular, symptom severity and duration often are, in effect, a single compound attribute with two dimensions. In such a case as this, respondents cannot evaluate outcomes where severity and duration are treated as separate attributes. For instance, respondents cannot assess migraine pain severity without knowing how long the pain will last and cannot assess migraine duration without knowing how severe the pain is during a specified period. Because the statistical model requires including an interaction between symptom severity and duration, the experimental design must ensure that it is possible to efficiently estimate such a model.

Cognitive limitations of particular groups of respondents

Because choice questions are cognitively challenging, statistically efficient designs may be beyond the reach of certain respondents, such as respondents with a condition that involves cognitive deficits including Alzheimer's disease, schizophrenia, or other neurological conditions. In such studies, the balance between acceptable response efficiency and statistical efficiency may have to favor simpler designs that yield less statistical information for a given sample size.

Labeled and constant alternatives

The majority of DCE studies in health care have used experimental designs with generic choice alternatives (e.g., medicine A, medicine B). Choice alternatives also can be given labels, where the alternative-specific label itself has some meaning or value apart from the specified attributes (e.g., nurse practitioner, general practitioner). Examples of this approach include Viney et al. [12] and Lancsar [33]. L^MA designs incorporate such labeled alternatives and allow for the independent estimation of alternative-specific attribute effects [20].

L^MA designs can be created by following standard approaches used to create generic designs (as discussed in the following sections), with the difference that alternative-specific attributes are treated as separate design columns. For example, when considering the choice of a health care provider, the alternatives labeled nurse practitioner and general practitioner can have separate parameter effects for an attribute such as waiting time. A useful feature of L^MA designs that use labeled alternatives is that they simultaneously create both the alternatives and the choice questions.

A related case is the presence of a constant alternative that has unchanging attribute levels in all choice questions. This alternative may describe a reference condition, the status quo, or an option to not participate (opt out). The presence of such an alternative can affect measurements of statistical efficiency, and many software packages can accommodate them via internal options or through the ability to specify user-defined constraints [15,34–37].

Blocking

Often, an experimental design that is constructed prior to fielding will contain more choice questions than the researcher wishes to ask to each respondent. In these situations, the researcher will have to carefully consider *blocking* the experimental design. Blocks are partitions of the choice questions (usually equally sized) in the experimental design that contain a limited number of choice questions for each respondent. In practice, respondents are randomly assigned to a block and answer the choice questions in that block instead of the entire design.

For example, a researcher may consider an experimental design with 24 choice questions but partition the design randomly into two blocks so that each respondent will answer only 12 choice questions. Blocking promotes response efficiency by reducing the necessary cognitive effort for each respondent who completes the survey. Desirable statistical properties of the experimental design (e.g., no correlations among attribute levels), however, may not hold for individual blocks. Certain software packages can perform the blocking prior to fielding, with varying control over the properties of each block.

Deviations from Strict Orthogonality

Orthogonality is a desirable property of experimental designs that requires strictly independent variation of levels across attributes, in which each attribute level appears an equal number of times in combination with all other attribute levels. Balance is a related property that requires each level within an attribute to appear an equal number of times. For the purposes of this article, we will consider a design to be strictly orthogonal only if it is orthogonal and balanced. Lack of strict orthogonality does not preclude estimating parameters. While nonzero correlations among attribute levels should be avoided, if possible, it is useful to note that market or revealed-preference data nearly always are collinear to some degree.

In practice, designs that are nearly balanced and nearly orthogonal usually are still well identified [38]. As long as the collinearity is not severe, all the parameters of interest will be sufficiently identified and estimation is feasible. In fact, the precision and accuracy of parameters may be improved by imposing constraints on the design that improve response efficiency and increase the amount of useful preference information obtained from a design of given size. Practical designs thus may deviate from strict orthogonality because of constraints placed on implausible combinations, lack of balance, or repetition of particular attribute levels across a set of alternatives (overlap).

Constraints on implausible combinations

Because all attributes in orthogonal designs vary independently, implausible combinations or dominated alternatives (where all the levels of one alternative are unambiguously better than the levels of a second alternative) are likely to occur. Dominated alternatives can occur when attribute levels a priori are naturally ordered. An example of a dominated alternative is a pair of treatments in which one alternative has unambiguously worse levels of health outcomes than another.

As discussed in the previous section, it often is advisable to impose restrictions prohibiting implausible attribute-level combinations from appearing in the experimental design. Dominated alternatives yield no information on trade-off preferences because all respondents should pick the dominant alternative, regardless of their preferences. While responses to these alternatives offer a test for measuring respondents' attentiveness to the attribute levels and definitions, such tests can be incorporated systematically outside of the design. Constraints that exclude implausible combinations or dominated alternatives introduce some degree of correlation and level imbalance in the experimental design.

Balance

All levels of each attribute appear an equal number of times in a balanced design, and balance is a necessary condition for strict orthogonality. Balance requires that the total number of alternatives (the number of questions multiplied by the number of alternatives in each set) should be evenly divisible by the number of levels for each attribute. For example, if the design includes three-level and four-level attributes, the total number of alternatives must be divisible by both 3 and 4 to ensure balance (i.e., 12, 24, 36, etc.).

If the design includes both two-level and four-level attributes, the total number of alternatives must be divisible by 2 and 4 (i.e., 4, 8, 12, 16, 20, 24, etc.). Note, however, that even if each level within an attribute appears an equal number of times in the design, each level in a three-level attribute will appear in one-third of the profiles and each level in a four-level attribute will appear in only one-fourth of the profiles. Other things being equal, we thus would expect wider confidence intervals for attributes with a larger number of levels because there are fewer observations available for estimating each level parameter.

Overlap

An attribute is overlapped in a choice question when a set of alternatives has the same level for a given attribute. Overlap provides a means for simplifying choice questions by reducing the number of attribute differences respondents must evaluate. Overlap thus can improve response efficiency. Overlap, however, may reduce design efficiency for a given number of questions and sample size because it potentially limits the amount of trade-off information obtained by the design (although the researcher could add more choice questions to overcome this) [39]. Some design approaches preclude any overlaps in the design, some approaches result in a few overlaps as a result of the procedure used, and some approaches allow researchers to control the pattern of overlaps allowed in the design.

Design Approaches and Software Solutions

While several measures of statistical efficiency have been proposed, D-efficiency, or D-optimality, remains the most commonly used metric in design construction [40]. The D-optimality criterion minimizes the joint confidence sphere around the complete set of estimated model parameters by maximizing the determinant of the inverse of the variance-covariance matrix in maximum-likelihood estimation. Most available experimental-design software solutions use algorithms to construct D-optimal designs for the smallest possible design that identifies all the necessary parameters. They also provide a number to measure D-efficiency, known as a D-score.

D-efficiency can be reported either *absolutely* or *relatively*. Absolute or raw D-efficiency refers to the D-score for a given experimental design. The absolute D-score depends on the coding scheme, model specification, attribute levels, and the priors for model coefficients that are specified in the design construction. Alternatively, a relative D-score enables comparison of multiple designs within a class; it is the ratio of D-scores between the proposed experimental design and a comparator design. It is invariant under different coding schemes but still is dependent on the model specification, attribute levels, and priors [38]. Most software packages present both measures, and so the researcher must take care in using the appropriate metric consistently when evaluating prospective designs. The default in most packages is the relative measure, which may be the most useful to practitioners, although it can also be misleading when the comparator design is of a type quite different from the generated design. Although there is no established threshold for what constitutes a best-practice efficiency score, researchers should evaluate relative efficiency together with other experimental-design criteria in assessing alternative designs.

In addition, researchers may be most interested in estimating ratios of parameters to obtain, for example, willingness-to-pay estimates or other types of marginal trade-offs. For these studies, an optimal design might strive to minimize the variance of the particular ratio of interest, again minimizing the confidence interval for a given sample size.

Actually finding a D-efficient design requires identifying a subset of the full-choice design of all meaningful combinations of attribute-level combinations placed into groups of alternatives. By completely enumerating all possible choice questions, the full-choice design is usually perfectly orthogonal in both main effects and all possible interactions. The size of a full-choice design, however, usually is impractically large. Researchers thus must accept the compromises required in using a subset of the full-choice design. Because all possible choice questions cannot be used, empirically feasible choice designs support identifying only main effects and some interactions; however, some higher-order effects are necessarily confounded [21,22,38,41].

While full-choice designs are strictly orthogonal because they are balanced and each pair of attribute levels appears with the same frequency, the term "orthogonal" is generally applied to a small subset of the full-choice design. In an orthogonal main-effects plan (OMEP), all main effects are uncorrelated with each other. OMEPs are optimal for main-effects linear statistical models. Main effects, however, can be correlated with interactions, and interactions can be correlated with other interactions.

Despite the attractive statistical properties of OMEPs, researchers may find them to be intractable or inflexible. Furthermore, OMEPs may not even exist for most combinations of attributes, levels, and numbers of profiles. In these cases, researchers have to use iterative search procedures and algorithms in software packages to find a D-optimal design that satisfies study constraints. Conceptually, these techniques methodically scan subsets of the full-choice design and return a specified number of choice questions with the specified number of alternatives that satisfy specified design criteria and approximate maximum D-efficiency. That is, the procedures and algorithms provide near-maximum D-efficiency, given the assumptions imposed by the researcher.

The search for D-optimal designs is complicated by the information required to calculate the D-score measure of efficiency for a particular design. The D-score is based on the determinant of the variance-covariance matrix, which, in turn, depends on both the specification and the parameter values for nonlinear models. Experimental designs that incorporate informative priors thus can be statistically more efficient than designs that assume uninformative priors that all parameters are equal to zero. Researchers may have information about the relative sizes of parameters based on previous studies, pretest data, pilot-test data, or logic [10].

Such priors may also affect response efficiency by changing the likelihood that particular trade-offs are evaluated by respondents. Even if there are no previous data to inform expectations about relative sizes of effects, naturally ordered categorical attributes at least convey information about the order of attributes and help identify dominated pairs of choice alternatives. Applying incorrect priors, however, may degrade the expected efficiency of the experimental design relative to a design with uninformative priors [10]. More advanced design-construction approaches allow researchers to specify Bayesian distributions of possible parameter values [42] or through specifying multiple efficient designs that cover more of the design space [43].

Kanninen [22] offered a solution to this problem, using updated priors based on intermediate data sets. Her approach assumed that at least one attribute was continuous—in other words, rather than assuming, a priori, that attributes can be represented only by one of a few discrete levels, in fact at least one attribute can take any value. Price is an example of an attribute that can have this flexibility, at least within a certain range. By allowing for this continuous attribute, the D-optimal design problem becomes a basic calculus problem: maximizing a criterion (D-score) over a continuous variable (the continuous attribute). Kanninen [22] showed that D-optimal designs derived under this assumption could be completely defined as orthogonal arrays but with one attribute varying enough so that certain, specific choice probabilities were obtained. Bliemer and Rose [44] refer to these optimal choice probabilities as "magic Ps," a term defined by an earlier research group who independently obtained similar results [45,46].

For practical implementation, researchers should conduct a pretest or should interrupt data collection at some point, based on an initial design using any of the available approaches. After collecting preliminary data, researchers can estimate parameters and calculate the sample probabilities for each choice question in the design. Researchers then adjust the continuous attribute to move the sample probabilities in the next round of data collection closer to the optimal "magic Ps." For many practical designs, the choice probabilities for two-alternative questions should be approximately 0.75/0.25 [22].

Studies that involve interaction effects prohibit implausible combinations and dominated alternatives or use strategies such as overlaps to improve response efficiency. Such constraints require more complex and statistically less efficient experimental designs to ensure model identification. Thus, the search for an optimal design is best characterized as maximizing the D-score subject to available information on likely parameter values and various constraints to improve response efficiency and achieve other study objectives. Some design-construction approaches can accommodate such flexible specification of design features and constraints; other approaches have fewer capabilities and thus are more suitable for simpler research problems.

Thus, the term *optimal* is a highly qualified concept in experimental design. The complexity of the experimental-design problem inevitably leads to pragmatic compromises to find a design that allows satisfactory identification of an intended statistical model. We create designs that we know are not perfect, but these designs are good enough to identify the parameters of interest under particular simplifying assumptions. We also seek to optimize the design with respect to a specified index of statistical efficiency, given the practical limitations of empirical research.

Comparison of Design Approaches

A principal objective of all experimental-design approaches is to maximize statistical efficiency for a given model, subject to various assumptions and possible constraints. In addition, each approach has ancillary objectives, such as using an efficient algorithm to construct experimental designs or minimizing the amount of programming knowledge or set-up complexity required. Different design approaches have emphasized one or more of these objectives. Some approaches are more concept-driven and incorporate newly developed algorithms or design strategies, whereas other approaches use pragmatic strategies that provide researchers with flexible, inexpensive, and easy-to-use tools for constructing a particular kind of design.

Each approach also uses a particular coding format for categorical variables, which, as described in the previous sections, may affect the interpretation of efficiency measures. While dummy coding is commonly used in empirical research to estimate a separate effect for all but one level for a categorical variable, many DCE researchers advocate using effects coding [19]. Several popular software programs use effects coding to construct experimental designs. Some designs, however, are based on other, more complicated coding schemes, such as orthonormal coding or orthogonal-contrast coding. Each scheme may possess certain advantages or disadvantages, and the researcher should be aware of the utilized format when interpreting the choice-model parameter estimates.

Although the heterogeneity in experimental-design approaches and objectives defies simple classification, this section describes features of several experimental-design approaches that are accessible to most users. The following sections summarize the features of six approaches:

- Orthogonal designs that can be constructed without the assistance of special software (manual catalog-based designs);
- SAS (Cary, NC) experimental-design macros (SAS macros);
- Sawtooth Software (Orem, Utah) choice-based conjoint designs (Sawtooth Software);
- Street and Burgess's cyclical designs (Street and Burgess);
- Sándor and Wedel's Bayesian designs (Sándor and Wedel); and
- Bliemer and colleagues' generalized approach to experimental design (Bliemer et al.)

Later in this section, we summarize the features of each approach: the modeling assumptions required to estimate preference parameters; whether the approach can accommodate restrictions, such as implausible combinations or number and type of overlaps; the use of prior information on the size of preference parameters; the coding procedure used for the variables; and the usability, availability, and cost of software for each approach.

Manually Constructed Designs

Catalog, fold-over, and other do-it-yourself approaches involve manually constructed experimental designs often based on OMEPs. Designs based on OMEPs support independent estimation of main-effect parameters for linear statistical models. OMEP designs do not allow independent estimation of interactions among attributes. Researchers have tended to favor the use of OMEPs because these designs are the most parsimonious in terms of the numbers of alternatives and choice questions required to obtain identification of the main effects. These designs also exhibit the two desirable design properties of orthogonality and level balance. The increased availability of software that facilitates the construction of more complicated designs has resulted in fewer studies that rely on catalog-based designs.

OMEP profiles correspond to the first alternative in a choice question. Profiles of alternatives in each choice question are constructed by systematically manipulating attribute levels, using one of a number of strategies, including fold-over, rotated, or shifted-design techniques. Some early researchers simply randomly combined pairs from the OMEP, but such an approach is likely to be highly inefficient. The fold-over approach replaces each attribute level with its opposite. For example, if there are two levels, L, for each attribute k (where $L_k = 2$) and a profile (with four attributes) is coded 0110, the fold over is 1001. If $L = 3$ and a profile (with six attributes) is coded 110022, it would be paired with 112200. Fold-over designs are orthogonal, but this approach is limited when $L_k > 2$.

Rotated designs create profiles of alternatives in each choice question by rotating each attribute level one place to the right or by wrapping around to the start of the sequence. A design that rotates the attribute levels would convert the profile 0123 to 1230. Rotated designs exhibit minimal level overlap, balance, and orthogonality but are restrictive because every choice question contains the same incremental difference.

Shifted designs use a generator and modular arithmetic (mod L_k) to create alternatives in each choice question. For example, to create an alternative for the profile 2102, modulo 3 arithmetic and the generator 1212 could be used to generate the profile 0011. Shifted designs exhibit orthogonality and minimal level overlap. OMEP-based designs do not allow imposing constraints on implausible combinations, dominated pairs, or overlap.

OMEPs can easily be obtained from design catalogs such as Hahn and Shapiro [47], from software packages including Orthoplan [48] and MktEx implemented in SAS 9.2 [38], or from online tables of orthogonal arrays [49,50]. Catalog and online OMEPs are available without licensing. Orthoplan is included in the SPSS basic package, and the MktEx macro is free to use within the SAS platform. Generating the profiles of alternatives for attributes with more than two levels can be complex to implement without software [38].

SAS Macros

Most researchers do not construct choice designs by direct means. They generally rely on procedures that use a computerized search algorithm. Adaptations of an algorithm first proposed by Fedorov [21,41,51–53] are well suited for this problem. The SAS system offers a variety of experimental-design macros that implement this approach. The algorithm typically starts with a random selection from a candidate set of profiles. The candidate set of profiles can be an array, an OMEP, or a nearly orthogonal design that incorporates user-specified constraints. Macros available in all standard installations of the SAS System allow researchers to select an orthogonal array from a preprogrammed library, to directly create an orthogonal array, or to construct a nearly orthogonal design. The SAS macros also allow for flexible constraints on the design.

Beginning with the first profile, the algorithm systematically exchanges a profile with another profile from the candidate set and determines whether the swap increases D-efficiency or violates any constraints. The algorithm then proceeds to the next profile and makes exchanges that increase D-efficiency until specified convergence criteria (size of the improvement, maximum time, number of iterations, etc.) are met. Our experience indicates that the algorithm converges to small improvements in the D-score rather quickly.

The SAS macros are well documented and provide numerous examples of how to construct designs for a wide range of applications [38]. If the experimental design is relatively simple, researchers with basic proficiency in SAS programming can generate an efficient choice design with little effort. The macros allow a variety of user-specified options, such as restricting duplicate profiles, blocking the design into versions that show a limited number of choice questions to each respondent, presenting status-quo alternatives, and using different coding schemes. If users require designs for more complicated models and profile constraints, they will need some proficiency in programming SAS macros, loops, and other procedures.

The SAS macros require access to a basic installation of the SAS System, thus requiring researchers or their organizations to purchase a SAS license. The macros themselves, along with extensive documentation on experimental-design methods, can be downloaded free of charge [38].

Sawtooth Software

Sawtooth Software supports several forms of conjoint analysis other than DCE, including adaptive conjoint analysis and adaptive choice-based conjoint analysis. In this article, we limited our comparison to the choice-based conjoint-analysis module (part of the SSI Web software platform). Unlike most other packages that create a fixed set of profiles by drawing from a subset of the full-choice design, Sawtooth Software's module samples from a subset of the full-choice design for each respondent while ensuring level balance and near-orthogonality within each respondent's profile. This approach avoids systematic correlations among interactions inherent in fixed designs and thus both main effects and higher-order interactions can be robustly estimated with sufficiently large sample sizes.

Sawtooth Software's approach can generate as many as 999 blocks of the design and assign each respondent randomly to a block. Sawtooth Software's procedure ensures that respondents see well-balanced and near-orthogonal fractions of the full-choice design. The procedure does not formally estimate D-efficiency and assumes that designs that are level balanced and near orthogonal will lead to identified preference-model parameters. Using a unique randomized design for each respondent reduces context effects. A disadvantage, however, is that design heterogeneity could be confounded with taste heterogeneity and scale differences.

Sawtooth Software provides users with several design options. The complete-enumeration procedure samples a subset of the full-choice design, with three additional modifications in mind: minimal overlap, level balance, and orthogonality. A shortcut scheme follows a procedure similar to complete enumeration, except that orthogonality is not strictly considered. Designs, however, are nearly orthogonal because of randomization. The software also includes a fully-random method that draws unique profiles with replacement from a subset of the full-choice design. Finally, the balanced-overlap method is a mixture of the complete enumeration and random methods. This procedure allows more overlaps than does the complete enumeration method but fewer overlaps than does the random method. Conditions for orthogonality are well controlled, and all options allow researchers to incorporate restrictions on implausible combinations.

Estimates of interaction effects in designs prepared by the choice-based conjoint-analysis module are unbiased; however, the efficiency of the estimate depends entirely on sample size. Because of the nature of randomized designs, all potential two-way interaction effects may be estimated with reasonable precision if sample sizes are sufficiently large. Thus, researchers do not need to identify specific interaction effects of interest at the outset of the study.

Set-up and management of design construction is handled through a simple, intuitive user interface. The software is designed to be accessible to users with a wide range of backgrounds and does not require programming skills. Using the Sawtooth Software program requires purchasing a software license [36]. The purchase includes the design software and full implementation of survey-instrument construction, administration, and analysis for choice-based conjoint-analysis or DCE studies.

Street and Burgess Designs

Street and Burgess [16] have developed a theory of the optimal efficiency properties of choice experiments in the logistic regression family. Indeed, Street and Burgess's designs are one of the few types of DCE designs available for which optimal efficiency properties are known (formal optimality properties are also known for designs that vary in price) [22]. Street and Burgess use this theoretical framework to produce optimal and near-optimal designs for generic, forced-choice, main-effects experiments for any number of alternatives and any number of attributes with any number of levels, assuming zero priors on the preference parameters and a conditional-logit model. The authors also provide a theory to construct choice experiments for main effects plus interactions if all attributes have two levels.

A key advantage of having formal proofs of optimality properties of DCE experimental designs is that the efficiency of any proposed design can be calculated relative to the conceptually most efficient design for a particular problem. Thus, the statistical efficiency of various designs in the logistic family can be compared. The approach uses orthonormal coding to achieve the calculated theoretical efficiency [54].

Street and Burgess's approach uses a shifting procedure applied to a starting design for the first profile in each choice question to create the other alternatives in each choice question. Generators are a set of numbers that are applied to the starting design to shift the levels on the attributes on the basis of orthogonal arrays, as described in the preceding sections. The number of attribute levels that vary across alternatives in each choice question is controlled by the way the generators are chosen.

Such designs are straightforward to construct by using the authors' free software. Users specify a starting design that can be created by hand or obtained from the authors' software page, from catalogs, or from other sources. Researchers then define the number of alternatives per choice question and other parameters; the software generates the choice questions. The software can be used to check the properties of designs obtained from other sources for identification and efficiency, given the assumptions of the underlying theoretical model.

Street and Burgess's main-effects designs tend to vary most or all attribute levels across alternatives in each choice question, thus in principle encouraging respondents to evaluate differences in all attributes in each choice question. The software does not allow constraints on implausible combinations or dominated pairs, nor does the software allow control of overlap patterns; however, limited control is available through the choice of starting design and selection of generators. The software and documentation are freely available on the authors' Web site [34]. The program is run on the Web site itself and does not require downloading any files. The software does not require users to have programming skills.

Sándor and Wedel Designs

Sándor and Wedel describe procedures to construct locally optimal experimental designs that maximize the D-efficiency of the linear multinomial logit model [42] or cross-sectional mixed multinomial logit models [43,55]. (Discussion of statistical modeling approaches is beyond the scope of this article; see cited references for details.) Users specify the number of alternatives for each choice question, whether or not a constant alternative is included, the priors for model coefficients, and the coding structure. The D-score is maximized by using heuristic procedures similar to the relabeling and swapping algorithms proposed by Huber and Zwerina [56], but Sándor and Wedel developed an additional cycling procedure to cover more of the design space. The authors have explored how misspecified priors affect design efficiency. The authors allow using a Bayesian prior distribution of parameter estimates to account for uncertainty about parameter values.

Most practical designs produce too many choice questions for a single respondent. Such designs are blocked into smaller sets of questions: data from the full design are collected from the sample, but individual respondents see only part of the design [20]. Sándor and Wedel [43] describe a procedure to ensure that blocked designs are jointly and locally optimal across blocks.

The Sándor and Wedel procedures are written by using the GAUSS [37] programming language. Consequently, the experimental-design code must be adjusted according to the requirements of each study. The procedures are flexible because a variety of requirements and assumptions can be incorporated into the existing program code as required. Considerable understanding of the statistical models, design algorithms, and the GAUSS programming language is required, which makes the implementation of the procedures more difficult when compared with experimental designs constructed by using other approaches.

Sándor and Wedel provide the experimental-design code on request but also recommend an advanced algorithm presented by Kessels et al. [57]. The code can be examined outside the GAUSS environment. Researchers can translate the GAUSS code into other programming languages, but the requirement of modifying the existing statistical code adds increased complexity.

Bliemer et al. Designs

Bliemer, Rose, and various collaborators have developed a number of extensions and generalizations of previous experimental-design technologies. By using the same general methodological framework used by Sándor and Wedel [42,43,55], Street and Burgess [16], Kanninen [22], Bliemer and Rose [29,44], and Bliemer et al. [30] derived a statistical measure for calculating the theoretical minimum sample-size requirements for DCE studies. They proposed the use of sample-size efficiency (S-efficiency), rather than D-efficiency, as the optimization criteria, subject to usual prior assumptions on parameter values.

In addition, Rose and Bliemer [58], Jager and Rose [59], Rose et al. [28], and Scarpa and Rose [60] have developed design procedures to account for several generalizations of the basic choice model, including procedures that account for modeling the effects of covariates in addition to attributes and levels, allow joint optimization of attribute levels, and allow for respondent-specific constant alternatives. Bliemer et al. [30] and Bliemer and Rose [44] also extended the efficient-design framework to include designs for nested-logit and panel mixed multinomial-logit models. Analogously to Sándor and Wedel's incorporation of Bayesian uncertainty about parameter distributions, Bliemer et al. [30] introduced a method to account for uncertainty as to what kind of model will be estimated once data have been collected.

These innovations have been incorporated in the Ngene software package [35]. Ngene allows users to generate designs for a wide variety of model specifications, including highly flexible specification of constraints and interaction effects. Additional features include optimization for a number of advanced choice-model specifications, Bayesian priors, D-efficient and other optimality criteria, willingness-to-pay optimization, alternative coding structures, and optimization algorithms. Ngene also includes various diagnostic tools to compare designs under alternative assumptions.

The Ngene software is based on syntax command structures similar to those used by the Nlogit module available as part of the Limdep software package. Design set-up and management are handled through an interface that requires some familiarity with program syntax conventions.

Ngene requires command syntax common to that of the Limdep and Nlogit software. The setup can be complex for more advanced experimental designs. The Ngene software, including the manual, may be downloaded for free [35,61]. Generating designs with the software, however, requires purchase of a license.

Conclusions

As outlined above, the strength of DCE methods is the ability of research to present stimuli to respondents in a controlled, experimental environment to quantify respondents' trade-off preferences. Traditional approaches to the development of experimental designs for DCE have focused on the relative statistical efficiency of such designs (i.e., identifying designs that get the most precise parameter estimates possible for a given sample size). This task force report emphasizes the overall efficiency of the experimental design—which depends on both statistical and response efficiencies. It is possible to create choice questions that have ideal statistical properties but which respondents cannot answer well or possibly cannot answer at all because of inherent contradictions in the outcomes, interventions, or technologies described. Some deviations from the statistical ideal may still result in satisfactory identification of the model parameters while actually yielding more precise estimates than could be obtained from a perfectly orthogonal design.

This report provides an overview of the role of experimental designs for the successful implementation of DCE methods in health care studies. Our article outlines the theoretical requirements for designs that identify choice-model preference parameters and summarizes and compares a number of available approaches for constructing experimental designs. We have not attempted to evaluate or endorse one approach over another. Rather, we have provided researchers with information to guide their selection of an approach that meets the particular requirements of their studies.

Several of these approaches are accessible to researchers at low cost or at no charge. Thus, well-constructed experimental designs are within the reach of both experienced stated-preference researchers and relative newcomers to this field of research. We encourage researchers to take advantage of recent theoretical developments and innovations in practical methods for design construction when developing efficient and effective experimental designs for DCE studies.

Several aspects of experimental design were outside the scope of this report. These include experimental designs for segmentation models and a review of the findings of the literature on "experiments on experiments." Also, while the list of design approaches discussed in this report includes the most common methods, the list is not exhaustive. Finally, experimental design is an area of active research. Nothing in this report should be construed as advocating limits on identifying and disseminating improved approaches to constructing better experimental designs.

Acknowledgments

The Task Force expresses gratitude foremost to Elizabeth Molsen, RN, Director of Scientific & Health Policy Initiatives, ISPOR, for organizing and managing this task force. Thanks also to the authors of the various design approaches for reviewing draft descriptions of their approaches, including Deborah Street, Leonie Burgess, Warren Kuhfeld, Bryan Orme, John Rose, Zsolt Sándor, and Michel Wedel. The Task Force also thanks reviewers who provided valuable comments in preparation of this manuscript, including Ben Craig, Brett Hauber, Jo Mauskopf, Christine Huttin, Sheryl Szienbach, Terry Flynn, Frank Liu, Kathy Beusterien, Bhagwan Aggarwal, Parul Gupta, and Ning Yan Gu. Any remaining errors are entirely the responsibility of the task force members.

Source of financial support: ISPOR provided support for a Task Force meeting held in Orlando, March 23, 2012.

REFERENCES

[1] Bridges JFP, Hauber AB, Marshall D, et al. Conjoint analysis applications in health—a checklist: a report of the ISPOR Good Research Practices for Conjoint Analysis Task Force. Value Health 2011;14:403–13.
[2] Johnson FR, Mansfield C. Survey-design and analytical strategies for better healthcare stated-choice studies. Patient 2008;1:299–307.
[3] Bridges J. Stated preference methods in health care evaluation: an emerging methodological paradigm in health economics. Appl Health Econ Health Policy 2003;2:213–24.
[4] Bridges J, Onukwugha E, Johnson FR, et al. Patient preference methods—a patient centered evaluation paradigm. ISPOR CONNECTIONS 2007;13:4–7.
[5] Szeinbach SL, Harpe SE, Flynn T, et al. Understanding conjoint analysis applications in health. ISPOR CONNECTIONS 2011;17:8–11.
[6] Marshall DBJ, Hauber AB, Cameron R, et al. Conjoint analysis applications in health—how are studies being designed and reported? An update on current practice in the published literature between 2005 and 2008. Patient 2010;3:249–56.
[7] Marshall DA, Johnson FR, Kulin NA, et al. Comparison of physician and patient preferences for colorectal cancer screening using a discrete choice experiment. Value Health 2007;10:A346.
[8] Orme BK. Getting Started with Conjoint Analysis: Strategies for Product Design and Pricing Research (2nd ed.). Madison: Research Publishers LLC, 2010.
[9] Ryan M, Gerard K. Using discrete choice experiments to value health care programmes: current practice and future research reflections. Appl Health Econ Health Policy 2003;2:55–64.

[10] Carlsson F, Martinsson P. Design techniques for stated preference methods in health economics. Health Econ 2003;12:281–94.
[11] Louviere JJ, Lancsar E. Choice experiments in health: the good, the bad, the ugly and toward a brighter future. Health Econ Policy Law 2009;4:527–46.
[12] Viney R, Savage E, Louviere J. Empirical investigation of experimental design properties of discrete choice experiments in health care. Health Econ 2005;14:349–62.
[13] Street DJ, Burgess LB, Viney RC, et al. Designing discrete choice experiments for health care. In: Ryan M, Gerard K, Amaya-Amaya M, eds. Using Discrete Choice Experiments to Value Health and Health Care. Dordrecht: Springer, 2008.
[14] Louviere JJ, Flynn T, Carson R. Discrete choice experiments are not conjoint analysis. J Choice Model 2010;3:57–72.
[15] Kuhfeld WF. Experimental design: efficiency, coding, and choice designs. 2010. Available from: http://support.sas.com/techsup/technote/mr2010c.pdf. [Accessed April 25, 2012].
[16] Street DJ, Burgess L. The Construction of Optimal Stated Choice Experiments: Theory and Methods. Hoboken: Wiley, 2007.
[17] Lancsar E, Louviere J. Deleting 'irrational' responses from discrete choice experiments: a case of investigating or imposing preferences? Health Econ 2006;15:797–811.
[18] Lancsar E, Louviere J. Conducting discrete choice experiments to inform healthcare decision making: a user's guide. Pharmacoeconomics 2008;26:661–77.
[19] Hensher DA, Rose JM, Greene WH. Applied Choice Analysis: A Primer. Cambridge: Cambridge University Press, 2005.
[20] Louviere J, Swait J, Hensher D. Stated Choice Methods: Analysis and Application. Cambridge: Cambridge University Press, 2000.
[21] Johnson FR, Kanninen B, Bingham M. Experimental design for stated-choice studies. In: Kanninen B, ed. Valuing Environmental Amenities Using Stated Choice Studies: A Common Sense Approach to Theory and Practice. Dordrecht: Springer; 2006.
[22] Kanninen B. Optimal design for multinomial choice experiments. J Mark Res 2002;39:214–27.
[23] Raghavarao D, Wiley JB, Chitturi P. Choice-Based Conjoint Analysis. Models and Designs (1st ed.). Boca Raton: Taylor and Frances Group, 2011.
[24] Finn A, Louviere JJ. Determining the appropriate response to evidence of public concern: the case of food safety. J Public Policy Mark 1992;11:12–25.
[25] Flynn TN. Valuing citizen and patient preferences in health: recent developments in three types of best-worst scaling. Expert Rev Pharmacoecon Outcomes Res 2010;10:259–67.
[26] 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;26:171–89.
[27] Louviere JJ, Flynn TN. Using best-worst scaling choice experiments to measure public perceptions and preferences for healthcare reform in Australia. Patient 2010;3:275–83.
[28] Rose JM, Bliemer MCJ, Hensher DA, et al. Designing efficient stated choice experiments in the presence of reference alternatives. Transp Res Part B Methodol 2008;42:395–406.
[29] Bliemer MCJ, Rose JM. Efficiency and sample size requirements for stated choice studies. 2005. Available from: http://ws.econ.usyd.edu.au/itls/wp-archive/itls_wp_05-08.pdf. [Accessed April 25, 2012].
[30] Bliemer MCJ, Rose JM, Hensher DA. Efficient stated choice experiments for estimating nested logit models. Transp Res Part B Methodol 2009;43:19–35.
[31] Maddala T, Phillips KA, Johnson FR. An experiment on simplifying conjoint analysis designs for measuring preferences. Health Econ 2003;12:1035–47.
[32] Johnson FR, Yang J-C, Mohamed AF. In defense of imperfect experimental designs: statistical efficiency and measurement error in choice-format conjoint analysis. In: Proceedings of 2012 Sawtooth Software Conference. Orlando, FL, March 2012.
[33] Lancsar E. Discrete choice experimental design for alternative specific choice models: an application exploring preferences for drinking water [Invited paper]. Presented at: the Design of Experiments in Healthcare Conference, Isaac Newton Institute of Mathematical Science, Cambridge, United Kingdom, 2011.
[34] Discrete Choice Experiments [computer software]. Sydney: School of Mathematical Sciences, University of Technology, 2007. Available from: http://crsu.science.uts.edu.au/choice. [Accessed April 20, 2012].
[35] Rose JM, Bliemer MCJ. Ngene. Available from: http://www.choice-metrics. com/download.html. [Accessed April 25, 2012].
[36] Sawtooth Software [computer software]. Sequim: Sawtooth Software, 2009.
[37] Gauss [computer software]. Gauss 8.0. Black Diamond: Aptech Systems, Inc., 2011.
[38] Kuhfeld WF. Marketing research methods in SAS. 2010. Available from: http://support.sas.com/techsup/technote/mr2010.pdf. [Accessed April 25, 2012].
[39] Sawtooth Software. On Interaction Effects and CBC Designs. Sequim: Sawtooth Solutions, 1998. Available from: www.sawtoothsoftware.com/education/ss/ss8.shtml. [Accessed April 20, 2012].
[40] Hall J, Kenny P, King M, et al. Using stated preference discrete choice modeling to evaluate the introduction of varicella vaccination. Health Econ 2001;11:457–65.
[41] Kuhfeld WF, Garratt M, Tobias RD. Efficient experimental design with marketing research applications. J Mark Res 1994;31:545–57.
[42] Sándor Z, Wedel M. Designing conjoint choice experiments using managers' prior beliefs. J Mark Res 2001;38:430–44.
[43] Sándor Z, Wedel M. Heterogeneous conjoint choice designs. J Mark Res 2005;42:210–8.
[44] Bliemer MCJ, Rose JM. Construction of experimental designs for mixed logit models allowing for correlation across choice observations. Transp Res Part B Methodol 2010;44:720–34.
[45] Fowkes AS, Wardman M. The design of stated preference travel choice experiments. J Transport Econ Policy 1988;13:27–44.
[46] Fowkes AS, Wardman M, Holden DGP. Non-orthogonal stated preference design. Proceed PTRC Summer Ann Meet 1993:91–7.
[47] Hahn GJ, Shapiro SS. Statistical Models in Engineering. New York: Wiley, 1967.
[48] SPSS [computer software]. Chicago: SPSS, 2008.
[49] Sloane NJ. A library of orthogonal arrays. 2003. Available from: http://www2.research.att.com/~njas/oadir/. [Accessed April 25, 2012].
[50] Kuhfeld W.F. Orthogonal arrays. Available from: http://support.sas.com/techsup/technote/ts723.html. [Accessed April 20, 2012].
[51] Cook RD, Nachtsheim CJ. A comparison of algorithms for constructing exact d-optimal designs. Technometrics 1980;22:315–24.
[52] Fedorov VV. Theory of Optimal Experiments. New York: Academic Press, 1972.
[53] Zwerina K, Huber J, Kuhfeld W. A general method for constructing efficient choice designs. Available from: http://support.sas.com/techsup/technote/mr2010e.pdf. [Accessed April 25, 2012].
[54] Rose JM, Bliemer MCJ, Stated choice experimental design theory: the who, the what and the why. In: Podium presentation presented at the 3rd Conjoint Analysis in Health Conference. Newport Beach, CA, October 5–8, 2010.
[55] Sándor Z, Wedel M. Profile construction in experimental choice designs for mixed logit models. Mark Sci 2002;21:455–75.
[56] Huber J, Zwerina KB. The importance of utility balance in efficient choice designs. J Mark Res 1996;33:307–17.
[57] Kessels R, Jones B, Goos P, et al. An efficient algorithm for constructing Bayesian optimal choice designs. J Bus Econ Stat 2009;27:279–91.
[58] Rose JM, Bliemer MCJ, Designing efficient data for stated choice experiments [poster]. Presented at: the 11th International Conference on Travel Behaviour Research. Kyoto, Japan, August 16–20, 2006.
[59] Jaeger SR, Rose JM. Stated choice experimentation, contextual influences and food choice: a case study. Food Qual Prefer 2008;19:539–64.
[60] Scarpa R, Rose JM. Design efficiency for non-market valuation with choice modelling: how to measure it, what to report and why. Aust J Agricult Res Econ 2008;52:253–82.
[61] ChoiceMetrics. Ngene 1.1.1 User Manual & Reference Guide, Australia; 2012.