

**A CHECKLIST FOR CONJOINT ANALYSIS APPLICATIONS IN HEALTH: REPORT OF THE
ISPOR CONJOINT ANALYSIS GOOD RESEARCH PRACTICES TASK FORCE
FINAL DRAFT: MAY 16, 2008**

For comment only.

Prior to quotation, please contact John Bridges. jbridges@jhsph.edu

John F. P. Bridges PhD, Assistant Professor, Department of Health Policy & Management, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, USA

Brett Hauber PhD, Senior Economist & Global Head, Health Preference Assessment Group, RTI Health Solutions, Research Triangle Park, NC, USA

Deborah Marshall PhD, Canada Research Chair, Health Services and Systems Research & Associate Professor, Department of Community Health Sciences Faculty of Medicine, University of Calgary, Calgary, Alberta, Canada

Andrew Lloyd DPhil, Director, Oxford Outcomes, Oxford, UK

Lisa A. Prosser PhD, Research Associate Professor, Child Health Evaluation and Research Unit, University of Michigan, Ann Arbor, MI, USA

Dean A Regier PhD, Senior Researcher in Health Economics, National Perinatal Epidemiology Unit, University of Oxford, Oxford, UK

F. Reed Johnson PhD, Senior Fellow and Principal Economist, Health Preference Assessment Group, RTI Health Solutions, Research Triangle Park, NC, USA

Josephine Mauskopf PhD, Vice President, Health Economics, RTI Health Solutions, RTI International, Research Triangle Park, NC, USA

Abstract

Background: The application of conjoint analysis methods (a general term spanning discrete choice experiments and other stated-preference methods) in health has increased rapidly over the past decade, yet broad acceptance of these methods is hindered by the variation in methods, terminology, and quality across these applications due to the lack of any clear and consistent methodological standards.

Objective: The ISPOR Conjoint Analysis Good Research Practices Task Force was established to identify good research practices in the application of conjoint analysis in health. In this paper the Task Force presents a conceptualization of, and guidance for, good research practices for authors, reviews, and readers of conjoint analysis applications in health and medicine.

Methods: The Task Force met regularly over the course of 24 months to identify the important stages of a conjoint analysis, to discuss good research practices for conjoint analysis, and to develop and refine the key criteria for identifying good research practices. ISPOR members contributed to this process through participation in multiple sessions at ISPOR conferences, presentations at the first and second Conjoint Analysis in Health Conferences, and via the ISPOR web site. The final set of good research practices is presented in the form of a checklist illustrated with stylized examples or with practical examples from the literature.

Results: Task Force findings are presented as a ten-item checklist, modeled on the iconic checklist by Drummond and colleagues for cost-effectiveness analysis, covering: i. Research question; ii. Attributes and levels; iii. Construction of tasks; iv. Experimental design; v. Preference elicitation; vi. Instrument design; vii. Data collection plan; viii. Statistical analyses; ix. Results and conclusions; and x. Study presentation.

Conclusions: The conjoint analysis checklist provides a framework for both assessing and improving the quality of conjoint analysis applications in health and medicine. Future work of the ISPOR Conjoint Analysis Good Research Practices Task Force will focus on specific guidance for the design, analysis, and interpretation of conjoint analysis applications in health. We will also focus on the role of conjoint analysis and related methods for identifying and valuing patient-relevant outcomes and focus on standardizing and documenting the vast array of terms used in the application of conjoint analysis.

Introduction

Purpose of this Article

Conjoint analysis is a stated preference survey method that increasingly is being used in health and medicine as a means to identify and value aspects of both health and health care (1-2). Given that conjoint analysis is grounded in several disciplines, including psychology, economics, decision sciences and marketing, there exists an abundance of different techniques and approaches to the method (3). In health and medicine, conjoint analysis has become an umbrella term that spans several different methods. Applications in health and medicine often refer to discrete choice experiments, stated-preference methods, best-worst scaling and many other terms. While many of these terms are used correctly, the variety of terms often makes such applications difficult to identify in the literature and difficult for a general medical or outcomes research audience to interpret.

Despite the growing number of applications of conjoint analysis in health, the broad acceptance of conjoint analysis methods is hindered by the variation in methods, terminology, and quality of their applications in these fields (4). Furthermore, the lack of methodological standards for the application of conjoint analysis in health and medicine likely has slowed the pace of publication of conjoint analyses. Likewise, variation in peer review and editorial processes among scientific journals has resulted in wide variation in the quality of published conjoint studies.

The ISPOR Conjoint Analysis in Health Good Research Practices Task Force was established to identify good research practices in the application of conjoint analysis in health and medicine. It is important to note that the members of the Task Force, and the members of the Patient Reported Outcomes & Preferences Special Interest Group's Patient Preference Methods (PPM) – Conjoint Analysis Working Group that oversee and initiated the agenda of the working group, agreed that the development of strict guidelines to conducting conjoint analysis in health and medicine was not possible for the following reasons:

1. One of the strengths of conjoint analysis is its flexibility, and the imposition of strict guidelines could limit its application.
2. Given the interdisciplinary nature of conjoint analysis, it would be unlikely that a consensus of best practices would be reached.

Many of the possible applications of conjoint analysis in health and medicine have been unexplored, underexplored, or under-published, so best practices would be difficult to establish during this current period of rapid methodological advancement.

As such, the Task Force endeavored to provide broad guidance on good research practices by offering some structure to the development, analysis, and publication of conjoint analyses in health. Therefore, this report deviates slightly from the traditional approach taken in ISPOR Task Force reports (5-7), in that it serves as part tutorial on conjoint analysis and part guidance on good research practices. The primary objective of this report is to present a checklist for identifying good research practices in the application of conjoint analysis in health and medicine suitable for researchers, peer reviewers, and users to improve the quality of published conjoint analyses

in these applications. The aim of the checklist is to provide a framework for assessing and improving the quality of research methods broadly defined as conjoint analysis, without favoring one technique over another.

Why are Preferences of Patients and other Stakeholders Important?

Understanding how patients and other stakeholders perceive and value different aspects of their health or of health-care interventions is vital to the optimal design and evaluation of programs. Incorporating these values in decision making ultimately may result in clinical, licensing, reimbursement, and policy decisions that better reflect the preferences of stakeholders, especially patients. Furthermore, aligning clinical practice, drug development, and health policy with patient preferences ultimately could improve the effectiveness of health interventions, possibly improving the adoption of, satisfaction with, and adherence to clinical treatments or public health programs.

What are Stated Preference Methods?

Stated-preference studies employ a variety of methods that are grounded in economic theory. These methods fall into two broad categories:

1. methods using ranking, rating, or choice designed to quantify preferences for various attributes of an intervention (often referred to as conjoint analysis, discrete-choice experiments, or stated-choice methods) and
2. methods using direct elicitation to estimate the monetary value of a single intervention (including methods known as contingent valuation, willingness-to-pay, and willingness to accept) (8-9).

The former methods are based on the economic theory of preferences and utility, while the latter methods are based on demand theory. Economic theories of preferences and demand are interconnected and, as such, these two approaches to stated preferences often overlap. For the purposes of this paper, we will focus on the former, and for simplicity, we will use the term conjoint analysis to describe the various related methodologies.

Conjoint Analysis in Health and Medicine

Conjoint analysis is used to measure the relative value of specific components of health status and health-care alternatives by decomposing an alternative into its constituent parts (10-13). For example, the component attributes that define a pharmaceutical intervention might include efficacy outcomes, safety and tolerability outcomes, mode of administration, and cost. In all conjoint analyses, different levels are assigned to each component attribute to create a series of profiles which study subjects are asked to evaluate through rating, ranking, or choice tasks. Subjects' systematic evaluation of these profiles allows researchers to infer the relative importance of each component attribute as well as changes in the levels of each component attribute. In this report, we define a conjoint analysis as any study in which subjects are asked to evaluate at least two profiles, each defined by at least two component attributes with at least two possible levels for each attribute.

Stated preference methods such as conjoint analysis are particularly useful for quantifying preferences for products or outcomes in cases where no markets exist or where market choices are influenced by regulatory and

institutional factors, such as in health and medicine (14). Conjoint analysis has been applied successfully to measuring preferences for a diverse range of health applications. Examples range from cancer treatments (15), HIV testing (16) and treatment (17), dermatology services (18), asthma medications (19), genetic counseling (20), weight-loss programs (21), insulin therapy in type 2 diabetes (22), diabetes prevention programs (23), colorectal cancer screening (24), and treatments for Alzheimer’s disease (25).

Developing the Checklist

The ISPOR Conjoint Analysis in Health Good Research Practices Task Force met regularly over the course of 24 months to identify the important stages of a conjoint analysis, to discuss good research practices for conjoint analysis, and to develop and refine the final checklist. ISPOR members were engaged in the development of the checklist through multiple sessions at ISPOR conferences (both at the International Meetings and European Congress), presentations at the first and second Conjoint Analysis in Health Conferences, and via the ISPOR web site. The final format of the checklist was then modeled after the ten-point checklist for cost-effectiveness analysis developed by Drummond and colleagues (26). Each item on the checklist is supplemented with key references, potential pitfalls, and recommendations for good research practices.

How should the checklist be used?

This checklist should be used to aid those new to conjoint analysis to understand the stages involved in producing good research. It also will be of benefit to those reviewing, reading, or otherwise assessing the validity of an application of conjoint analysis in health and medicine. By outlining a systematic process of good research practices for applying conjoint analysis – from formulating the research question through the presentation of the research (either in presentation, abstract, or manuscript) – we aim to facilitate the research process and to highlight important issues that often are neglected or poorly executed. In producing the checklist, we aimed to be as inclusive as possible – both in term of the accessibility to a broad readership and applicability to a variety of conjoint analysis methods. We have aimed at highlighting “good research practices” rather than “best research practices”. Hence we have written in a non-technical way, with many elements of the checklist presenting “food-for-thought” rather than a necessary or sufficient argument for research excellence.

We caution readers, and especially peer reviewers, that given the variety of conjoint analysis methods, a single standard of best practice is likely not to exist for many aspects of conjoint analysis and a pluralistic approach to review is needed. Throughout this report we encourage researchers to provide and reviewers to expect explanation and justification of the methods chosen to complete any given study, rather than promote any specific method. Like any contribution in health and medicine, the goal of good scientific practice is possible replication of the results. Thus, it often is more important to know what was done, rather than why it was done. While we do not advocate that the checklist be used to score articles – we specifically did not argue for a scoring method for the checklist or any threshold of acceptability – we acknowledge that it might be used in that way. Our hope is, however, that the checklist is used as means to promote more high-quality applications and subsequent publications of conjoint analyses in health and medicine, and not as a hurdle to publishing.

Description of the Checklist

The following checklist is designed to provide a guide for evaluating conjoint studies that can be used by conjoint researchers in conducting their own work and by readers and users of conjoint studies. The ten-point checklist is presented in Table 1. In the remaining sections, we describe issues to be considered in evaluating each of these ten items and provide references for researchers to consult to better understand these issues.

i. Research question

The first item in the checklist is common to all scientific research and relates to the research question. Specifically ***“Was a well-defined research question stated and is conjoint analysis an appropriate method for answering it?”***

Following generally accepted research practices, a conjoint analysis study must clearly state a well-defined research question that defines what the study will attempt to measure (8). For example, a conjoint analysis might be undertaken to quantify patients’ preferences for cost, risk of complications, and service location of a health care service. In addition to defining the research question, researchers should indicate the hypothesis(es) to be tested in the study. The hypothesis tests may be implicit in the research question itself. For example, in a conjoint analysis estimating the rate at which subjects are willing to trade off between two attributes, the testable null hypothesis is that the preference weight for one level of the attribute is not statistically significantly different from the preference weight for a different level of that attribute.

In other words, the hypothesis test is designed to infer whether a change in the level of one attribute (for example, a change in surgical wait time from one month to two months) is statistically significant. If the null hypothesis is rejected for a given attribute, then the change in the attribute level is statistically significant indicating that subjects are not willing to trade other attributes for changes in that attribute.

Second, the research question should define the study perspective including any relevant decision-making or policy context. The research question, “What are patients willing to pay for treatment to reduce the rate of relapse in multiple sclerosis?” includes both the items to be measured – the tradeoff between cost and reduction in relapse rates – the perspective and decision-context of the analysis – and the study perspective in the context of the patients’ decision about multiple sclerosis treatment.

In a health policy context, the research question might look something like, “What level of increase in waiting time for non-emergency surgery is the public willing to accept to reduce the rate of surgical errors?” Here, the items to be measured include waiting periods for non-emergency surgery and the rate of surgical errors. The perspective is that of the government or health service provider in which the preferences of the general public are estimated in the context of health policy for the delivery of non-emergency surgical health care services.

A conjoint analysis study should include reasons to justify that conjoint methods are appropriate to answer the research question. Conjoint analysis studies are well suited to the valuation of services or products that differ in their component attributes and where decisions-makers are willing to trade off among these component attributes. Because the examples of research questions presented above involve explicit tradeoffs between measurable attributes, conjoint methods are appropriate in each case.

However, conjoint analysis may not be the only method for answering these research questions. In the first example, contingent valuation is a likely alternative for estimating patients' willingness to pay to reduce the rate of relapse in multiple sclerosis. In the second example, observational data from a pilot program for reducing surgical errors in which surgical waiting times are increased may provide enough information to answer the research question. Therefore, researchers should identify not only whether conjoint analysis can be used to answer the research question, but also why conjoint analysis is preferable to alternative methods.

ii. Attributes and levels

A central feature of a conjoint analysis is the combination of the attributes and levels and is addressed in the second key point in the checklist: "***Were the attributes and attribute levels supported by evidence?***"

The objective of conjoint analysis is to elicit preferences over the full range of attributes and attribute levels that define the profiles in the conjoint tasks. All attributes that potentially characterize the alternatives should be identified. In addition, the attribute levels should encompass the range that may be salient to subjects even if those levels are hypothetical or not feasible given current technology.

The identification of attributes should be supported by evidence regarding the potential range of preferences that people may have. Sources of evidence should include: literature reviews and other evidence regarding the impact of a disease and the nature of a health technology; clinical experts, and, perhaps most usefully, interviews or focus groups with individuals who represent the population from which study subjects likely will be drawn. The choice of whether focus groups or interviews should be used depends on many factors including the nature of the questions being asked and the types of people being included in the research. Simple thematic analysis is probably sufficient for guiding attribute selection. Such qualitative research will provide the basis for identifying the full set of attributes and possible levels that characterize the profiles in the preference space. Discussion with experts and further pilot testing with subjects can be used to narrow down the list of attributes if necessary.

The subset of attributes from the preference space that should be included in the conjoint tasks can be determined based on three criteria: relevance to the research question; relevance to the decision context; and whether attributes are potentially correlated. Attributes central to the research question or to the decision context must either be included in the conjoint tasks or held constant across all profiles in the conjoint tasks. For example, when eliciting patients' preferences for a surgical intervention, the efficacy of the intervention almost certainly is an important outcome.

However, if the research is designed to estimate patients' willingness to pay or willingness to wait to have a less-invasive surgical procedure, then it may make sense to exclude efficacy from the set of attributes and to inform participants that efficacy does not vary between the profiles. Similarly, it also is important to control for any potential attributes that are omitted from the conjoint tasks, but which correlate with attributes that are included in the conjoint tasks. In the US healthcare market, insurance coverage and out-of-pocket medical expenses for procedures are routine for many patients. Cost may be perceived as correlated with improvements in medical outcomes or access to advanced interventions. If cost is not included in such a study, it needs to be controlled for by holding it constant across profiles.

Attribute levels are usually combined into profiles in the conjoint tasks using an orthogonal design. However, if attributes are correlated, an orthogonal design could result in illogical combinations of attribute levels in a given profile in the conjoint task. For example, if one attribute defining a profile is the need for medication (no medication required or 2 tablets a day), it makes no sense to combine certain levels of this attribute with certain levels of an attribute describing the side effects of the medication (mild or severe) because some profiles would include a combination of the “no medication required” level of the need for medication attribute with the “severe” level of the side effect attribute.

iii. Construction of tasks

Conjoint tasks can be assembled in a number of different ways, and hence it is important to ask, “***Was the construction of the conjoint tasks appropriate?***”

First, each profile that subjects are asked to evaluate could include the full set of attributes included in the study (a full-profile task) or a subset of the attributes included in the study (a partial-profile task). Prior to constructing conjoint tasks with full profiles, researchers should determine, through qualitative research or quantitative pilot tests whether or not subjects can reasonably evaluate the full profiles or if they will employ simplifying heuristics such as focusing on only a few attributes while ignoring others when completing the conjoint tasks. If this happens, researchers learn nothing about subject preferences among the attributes that the subjects ignore and the importance of the attributes on which subjects focus likely will be overstated. If each conjoint task contains a partial profile, researchers must understand the effect of omitting some attributes in some conjoint tasks while omitting other attributes in different conjoint tasks. The way in which certain attributes are omitted can introduce biases in the results of the study.

The number of conjoint tasks often varies from study to study depending on the type and difficulty of each conjoint task. In some studies, subjects may be presented with a set of many alternative profiles and asked to order or rank the profiles from most preferred to least preferred. In this type of study, subjects often complete only one conjoint task. In other studies, profiles are grouped in sets of two or three and subjects are asked to rate, rank, or choose among these alternatives. Because each conjoint task includes only a small subset of the overall number of potential profiles, subjects often are asked to complete multiple conjoint tasks. In this case, researchers should justify both the number of profiles in each conjoint task and the number of conjoint tasks included in the data collection instrument.

iv. Experimental design

Experimental design is the process of systematically manipulating the attribute levels to create the profiles and conjoint tasks, hence it is important to ask, “***Was the choice of experimental design justified and evaluated?***”

The aim of an experimental design is to create a set of conjoint tasks that will yield as much statistical information as possible to estimate the parameters of the underlying preference model (usually preference weights for all attribute levels) with the most precision (27-28). There are several design properties that should be considered when choosing an experimental design in a conjoint study. A design is orthogonal if all effects can be estimated independently of all other effects (the effects are uncorrelated). A design is balanced when each level of an attribute is presented the same number of times across the set of conjoint tasks. Furthermore, efficiency is a

measure of the goodness of the design, and an efficient design has a small variance matrix. If a design is orthogonal and balanced, then it has optimum efficiency.

The experimental design will affect the estimable functional form of the model and its statistical efficiency (i.e., the standard error around the parameter estimates). Researchers therefore, should examine and test several design approaches. A full-factorial design in which every possible profile is presented will allow for the independent estimation of all main effects and interactions (i.e., all main effects and interactions are orthogonal). However, the number of profiles required by a full-factorial design may be too numerous for subjects to feasibly evaluate. Large designs can be blocked into multiple sub-designs to limit the number of tasks each respondent must complete. This blocking approach also allows for the independent identification of all parameters and interactions, but a loss of orthogonality may occur if a proportion of the sub-designs are not returned (13). Researchers should therefore ensure that each sub-design is randomly assigned to subjects.

The fractional factorial design is an alternative approach to decreasing the number of conjoint tasks required in the data collection instrument. Fractional factorial designs are orthogonal profiles constructed from a subset of the full factorial. These designs guarantee that all attribute main effects are independently estimable and they allow for the independent estimation of some attribute interactions if such interactions *are* defined by the researcher a priori (13).

Fractional-factorial designs typically are generated using published “catalog designs” (29-30) or statistical programs (e.g, SAS, SPSS). In choice-based conjoint analysis, the orthogonal arrays are used as “seed” profiles and choice alternatives are generated from the seed design using techniques that enforce the design criterion of orthogonality (zero correlation between attributes), level balance (each attribute level occurs with equal frequency), and minimal level overlap (each attribute level only appears once in a given choice) (27). These criteria provide good statistical efficiency for linear models, but for non-linear probabilistic models (e.g., the conditional logit model) designs that are more statistically efficient can be constructed by using design algorithms to search for an experimental design that minimizes a summary measure of the information matrix (e.g., the D-error criterion, (27, 31).

v. Preference Elicitation

Given that the aim of a conjoint analysis is to measure preferences, it is important to ask, “**Were preferences elicited credibly?**”

There are multiple question formats that can be used in preference-elicitation studies. Different elicitation formats provide responses to different questions. Therefore, researchers should ensure that the elicitation format used in a conjoint study is appropriate to answer the research questions the study is designed to answer. For example, the most appropriate elicitation format may differ for choice experiments that evaluate a new drug, health service, or public health program. In addition, data generated using different question formats will require different methods of statistical analysis.

In a discrete-choice experiment or forced-choice conjoint study, each conjoint task includes two profiles, and each profile is defined by a set of attribute levels between which subjects are asked to choose. Alternative question formats include ratings or rankings. Ratings and rankings provide more information about preferences

than choice formats. Ratings indicate intensity of preference for one profile relative to another, possibly including indifference, and rankings indicate the relative position of multiple alternative profiles relative to each other.

Conjoint tasks also can include an opt-out alternative. In these cases, the opt-out alternative allows subjects to choose standard-of-care, current treatment, or no treatment rather than the hypothetical alternatives included in a conjoint task. When subjects choose an opt-out alternative in a specific conjoint task, researchers learn nothing about subjects' relative preferences for the hypothetical alternatives presented in the task.

Therefore, while including an opt-out alternative often may provide a more realistic scenario for subjects to evaluate, it also introduces additional challenges in the design and analysis of the study. An alternative to including an opt-out alternative in each conjoint task is to include an opt-out alternative in a separate question following each conjoint task – subjects who select alternative A in a forced-choice question are offered A or opt out in a follow-up question. Including an opt-out question in a follow-up task increases the length and difficulty of the conjoint survey instrument, but provides the researcher with a more complete set of preference data.

vi. Instrument design

Conjoint data collection instruments are surveys and the development of a conjoint data collection instrument should follow good survey research principles. Therefore, it is important to ask, **“Was the data collection instrument designed appropriately?”**

It is important to elicit subject-specific health and demographic information to test for systematic differences in preferences based on these characteristics. Patients' health status may influence their willingness to pay in a systematic way and so may reduce the generalizability of the findings (32.).

Sample size calculations represent a challenge in conjoint analysis. Minimum sample size depends on a number of criteria including the question format, the complexity of the choice task, the desired precision of the results, and the need to conduct subgroup analyses (28). Researchers commonly apply rules of thumb such as that proposed by (33), which suggests that 300 observations per attribute level are required. Simulation techniques, which have been used in EQ-5D valuation work could potentially be used (34). Sample size estimation for conjoint analysis requires further work because it is an important criterion for grant awarding bodies and ethics committees.

Because conjoint tasks often are complex and cognitively burdensome to subjects, potential measurement error may be a serious concern in conjoint studies (35). Measurement error may be introduced by the order in which attributes are presented, the question order, or the number of attributes and levels. Work by Kjaer and colleagues (36), for example, suggests participants can show a differential sensitivity to price depending on where the cost attribute occurs in the profile. Although it is probably best to not randomise the order of attributes across conjoint tasks within a survey, some of these issues could be addressed through randomising the order of questions.

Finally, it is important to pilot the final questionnaire with respondents using both small cognitive debrief interviews (n~5-10 people) and also a larger quantitative pilot (n~30). The cognitive debrief will identify areas of misunderstanding or common errors and whether the survey is too lengthy. It will also test whether people

understand the instructions and whether participants feel the questions are appropriate. The larger pilot allows for consistency or rationality tests and can give estimates of coefficient size and direction.

vii. Data collection

Given that conjoint analysis is an empirical method, it is important to ask, **“Was the data collection plan appropriate?”**

Conjoint analysis surveys can be administered in many different ways including postal surveys using a paper-and-pencil survey instrument (22), non-mediated paper-and-pencil surveys completed at a finite set of study sites (37) electronic administration at a finite set of study sites using a PC (38), or electronic administration over the Internet (39). The complexity of most conjoint questions probably precludes the use of telephone-based data collection unless the survey instrument is mailed to subjects in advance. Interviewer-led administration of the survey may improve the quality of data because the interviewer can more fully explain the task including answering some questions (obviously without leading the participant in any way), and recognise if more explanation is needed. However interviewers may lead to bias because participants feel they need to produce socially desirable answers or they may believe that their answers could influence their future care. McColl & Fayers (40) provide an interesting review of administration methods. Participants should provide consent according to human subjects regulations, and they should also be made aware that the choices they are presented with are not necessarily choices that could be available to them from their doctor.

Subject recruitment must be consistent with the study objectives. Specifically, the sample of subjects must reflect those people whose preferences we aim to elicit and to whom we wish to generalise. For example, if we wish to quantify the preferences of women with metastatic breast cancer then the sample should be drawn from this population.

However, it is often difficult or costly to generate a sufficiently large sample of respondents from a narrowly defined population. Therefore, it may be reasonable to elicit preferences from not only a sample with a specific medical condition, but also from a sample of people at risk for the specific medical condition. Using the example above, it may be reasonable to include women with localized tumors of the breast who can reasonably conceive of developing more advanced breast cancer. Where this is done it needs to be made explicit and addressed in the analysis.

International studies present their own challenges to conjoint surveys. It is very important to ensure that appropriate translation and cultural adaptation of the survey takes place if it is used in multiple countries. Methods used in the development of other patient reported outcomes instrument should be sufficient (41). It may also be necessary to undertake additional pilot work in the new countries to ensure that the attributes, levels and conjoint tasks are acceptable. If a cost attribute is included then it is not sufficient to simply use exchange rates and instead we would recommend the use of purchasing power parities.

The design and conduct of conjoint studies should also consider the participant and whether there are any issues which would affect their ability to complete the survey. Conjoint studies can be cognitively complex to complete. Some patient groups who have known cognitive function problems – such as people with neurological diseases –

may not be able to complete the tasks. Where there is suspicion that disability may affect people then it makes sense to simplify the conjoint tasks as much as is reasonable.

It is important to describe the subject-specific health and demographic characteristics of the sample and compare these characteristics to the population to which you wish to generalise. Reviewers and readers inevitably will question whether you have inadvertently captured the views of more highly educated, more proactive patients (such as those who participate in patient advocacy groups) or patients with higher-than-average incomes.

viii. Statistical analyses

Conjoint analysis data and the modeling of preferences can require some complex statistical methods, hence it is vital to ask, **“Were statistical analyses and model estimation conducted appropriately?”**

There are several objectives of analyzing stated-preference data. First, one wants to estimate the strength of preference for the attributes and attribute levels included in the survey. One might also be interested in estimating how preferences vary by individual-subject characteristics. For policy analysis, one might also be interested in calculating how choice probabilities vary with changes in attributes or attribute levels, or in calculating secondary estimates of money equivalence (WTP) (42), risk equivalence (maximum acceptable risk (MAR)) (43), or time equivalence for various changes in attributes or attribute levels (44).

Theoretically valid and unbiased preference estimates depend on model specifications that are consistent with the underlying utility theory used to elicit preferences and with the particular features of the response and profile variables. Forced-choice conjoint analyses, discrete-choice experiments, and rating studies lend themselves to analysis using stochastic utility maximization theory. Rating or card-sort conjoint data often are analyzed using ordinary least squares or ordered-probit methods.

In many conjoint analyses, multiple responses are obtained from each subject. In these cases, researchers should ensure that the statistical analysis of the conjoint data account for within-subject correlation. Ignoring the fact that each subject provides multiple responses can result in biased preference estimates. Thus, researchers who estimate these models should test that the data being analyzed is consistent with the assumptions required for the model being employed.

Researchers must determine whether to model attribute levels as continuous or categorical. If attribute levels are specified as continuous, researchers must determine the appropriate functional form for each continuous variable. Categorical models avoid imposing any functional form on preference weights and provide a validity check on the correct ordering of naturally ordered attribute levels. In addition, researchers should determine whether categorical attribute levels are specified as dummy variables or effects-coded variables. When effects coding is used, zero corresponds to the mean effect for each attribute, rather than the combination of all the omitted categories, and the parameter for the omitted category is the negative sum of the included-category parameters. Hensher et al. (13) explain why effects coding is statistically superior for choice models.

Finally, researchers also should account for differences in preferences that arise from differences in individual characteristics such as age, income, education, and gender by interacting individual characteristics with attributes

included in the conjoint questions or by conducting split-sample analyses. Latent-class models allow the data to determine the optimal division of observations into groups with similar preferences (28).

ix. Results and conclusions

Those using conjoint analysis are often prone to making inferences and predictions that go beyond what the data and methods can support, and in outcomes research it is especially important to ask, **“Were the results and conclusions valid?”**

Evaluating the validity of results and conclusions requires consideration of the research question as well as other aspects of the design and analysis. The results should present the statistical findings in the context of the research question and should be presented in sufficient detail. The results should state which attributes/levels (and interaction terms, if relevant) included in the conjoint task were or were not significant and report uncertainty associated with estimates. Findings should be interpreted in the context of the choice being considered.

For example, in the multiple sclerosis example, the results could indicate that the rate of relapse was a significant attribute and a negative coefficient might imply higher rates of relapse were less preferred. If attributes/levels were found to be non-significant in the statistical analysis, these findings should also be clearly stated in the results. Results also should provide interpretation of the relative value of specific attributes, such as how the acceptable waiting time for non-emergency surgery varies with the rate of surgical errors (i.e., the marginal willingness-to-wait for a reduced rate(s) of surgical errors). Statistical uncertainty should be reported in a manner consistent with the type of model selected. If alternative model specifications were tested, the results of these alternative analyses should be described if not presented in full.

Limitations of the study and the potential effect(s) of these limitations on results should be clearly identified. Limitations can arise from selection of attributes and/or levels, such as simplifications adopted during survey development in order to generate a feasible design, possible correlation among selected attributes, and other design features such as the inclusion or exclusion of an opt-out option. Assumptions underlying the analytic approach may also affect interpretation of results and should be discussed. If the study population is not representative of the population, this may limit generalizability of findings and any extrapolation of results beyond the study population should be qualified and discussed.

The conclusion section should identify key findings of the study in the context of the original research question. A key element of any research study is to provide a relevant framework for interpreting the results: whether the results are consistent with or differ from existing studies in the literature and how this study extends existing research should be clearly identified and discussed.

x. Study presentation

Finally, when evaluating a conjoint analysis study, the researcher should ask, **“Was the study presented well and completely?”**

The importance and context of the study must be adequately motivated so as to answer the “so what” question. First, the key background literature should be cited to place the study in an appropriate clinical or health policy

context. The specific contribution of the paper in terms of innovative methods or an important application should be clearly stated in the introduction.

The text describing the study should be worded and structured appropriately for the target journal and audience. Journals vary in both the type of reviewers as well as the eventual readers, but in general, the use of jargon should be minimized, and acronyms and technical language (e.g. “importance weights” and “fractional factorial design”), should be clearly defined. A journal such as *Value in Health* has reviewers and readers who are familiar with conjoint analysis methodology, but this likely is not the case for a clinically focused journal. Further, since conjoint analysis is a relatively new area of research, the use of technical terms is not always consistent among authors. For example, the terminology used for the results might be “importance weights” or “preference weights”. Such inconsistencies are confusing to reviewers and readers. Therefore, definitions should be included for the technical terms used in the paper including the alternative technical terms.

Since there are no standardized rules for creating a conjoint survey and there is an enormous number of possible experimental designs, the methods and rationale for the conjoint analysis data collection instrument design must be adequately described, including the qualitative research conducted to identify the attributes and levels, the experimental design to create the conjoint tasks, and the methods used to analyze the results.

The matrix of attributes and levels and the final conjoint analysis data collection instrument need to be submitted for review along with the paper. It is not possible for a reviewer to provide a meaningful review of a conjoint paper without seeing the format and framing of the questions that generated the data. The properties of the experimental design should be described to provide a context for the strengths and limitations of the survey results. For example, if the experimental design does not allow interactions to be tested (a main effects design), this assumption should be clearly disclosed in describing the methods.

Finally, the discussion section should focus on both the innovative features of the paper and the implications of the results for the target audience. The unique contributions of the conjoint analysis study should be discussed and compared in the context of the current state of knowledge based on the published literature and health policy climate. However, as in all research, authors must be careful not to overstate the importance of their findings. Because conjoint analyses in health and medicine are published in different types of journals and may use different terminology, it is important for authors to ensure that what may appear to be novel has not been conducted previously. In addition, it is important that authors inform readers that the results of a conjoint analysis often provide estimates of the value or importance of attributes to subjects, but often do not, in and of themselves, predict future behavior or health outcomes.

The findings should be evaluated with respect to the research question that the study was designed to answer and the hypothesis(es) to be tested in the study. If the target audience is a clinical one, the conclusions of the paper should focus on the clinical implications of the study findings. For example, the results can be translated into simple statements about their possible impact on physician practice. Alternatively, if a study was designed to inform health policy, the findings about public, patient or provider preferences can be translated into suggestions for increasing the appropriate use of health care services. For example, in a conjoint analysis of colorectal cancer screening tests, the findings were translated into changes in rates of uptake of colorectal cancer screening based on the mix of alternative screening tests offered (45).

Conclusions

This report presents a checklist for good research practices for the application of conjoint analysis in health and medicine and is based on the consensus and international experience of the Task Force members. It is important to note that this consensus relates to the questions raised as part of the checklist and not the answers. Given that conjoint analysis is an extremely flexible tool and represents an interdisciplinary approach to understanding the preferences of patients and other stakeholders, it is unlikely that any consensus on methods will be reached. Such variation in methods is not unique to conjoint analysis; for example, there is a hearty debate concerning risk adjustment methods in outcomes research. Furthermore, unlike cost-effectiveness analysis where a global view of health care interventions is taken, conjoint analysis aims to be both specific and descriptive. Hence, we believe that a “reference case” for conjoint analysis is neither necessary nor likely to emerge.

Acknowledgements: This project was completed under the guidance of the ISPOR Patient Reported Outcomes & Patient Preferences Special Interest Group’s Patient Preference Methods (PPM) – Conjoint Analysis Working Group. We are grateful for the current and past members of this working group and associated reference group. We are particularly indebted to Elizabeth Molsen and Marilyn Dix Smith for challenging and empowering the Conjoint Analysis Task Force and PPM – CA working group to achieve and to broaden the methods available to outcomes researchers worldwide.

Source of financial support: None.

References:

1. Bridges, J, Kinter, E, Kidane, L, Heinzen, R, and McCormick, C (2008) “Things are looking up since we started listening to patients: Recent trends in the application of conjoint analysis in health 1970-2007” *The Patient – Patient Centered Outcomes Research*, 1(4), 273-282.
2. Ryan M, Gerard K. 2003. Using discrete choice experiments to value health care programmes: current practice and future research reflections. *Appl Health Econ Health Policy*, 2:55–64.
3. Hanley N, Ryan M, Wright R. Estimating the monetary value of health care: lessons from environmental economics. *Health Econ* 2003;12(1):3-16.
4. Johnson FR, Mansfield C. (2008) Survey-Design and Analytical Strategies for Better Healthcare Stated-Choice Studies. *The Patient – Patient Centered Outcomes Research*, 1(4), 299-308.
5. Wild D, Grove A, Martin M, Eremenco S, McElroy S, Verjee-Lorenz A, Erikson P; ISPOR Task Force for Translation and Cultural Adaptation. Principles of Good Practice for the Translation and Cultural Adaptation Process for Patient-Reported Outcomes (PRO) Measures: report of the ISPOR Task Force for Translation and Cultural Adaptation. *Value Health*. 2005 Mar-Apr;8(2):94-104
6. Mauskopf JA, Sullivan SD, Annemans L, et al. Principles of Good Practice for Budget Impact Analysis: Report of the ISPOR Task Force on Good Research Practices – Budget Impact Analysis. *Value in Health* 2007;10;336-47
7. Peterson AM, Nau DP, Cramer JA, et al. A checklist for medication compliance and persistence studies using retrospective databases. *Value Health* 2007;10(1):3-12
8. Bridges JFP. Stated-preference methods in health care evaluation: an emerging methodological paradigm in health economics. *Applied Health Economics and Health Policy* 2003;2:213-24.
9. Bridges J, Onukwugha E, Johnson FR, Hauber AB (2007) “Patient Preference Methods – A Patient Centered Evaluation Paradigm”, *ISPOR Connections*, 13(6): 4-7.

10. Louviere J, Hensher D, Swait J. Stated Choice Methods: analysis and applications. 1 ed. Cambridge: Cambridge University Press, 2002.
11. Viney R., Lanscar E., Louviere J. Discrete choice experiments to measure consumer preferences for health and healthcare. *Expert Rev Pharmacoeconomics Outcomes Res* 2002;2:89-96.
12. Lanscar E, Louviere J. Conducting discrete choice experiments to inform healthcare decision making: a user's guide. *Pharmacoeconomics* 2008; 26(8): 661-677.
13. Hensher DA, Rose JM, Greene WH. *Applied Choice Analysis: A Primer*. Cambridge, UK: Cambridge University Press, 2005.
14. Ryan M.,Farrar S. Using conjoint analysis to elicit preferences for health care. *BMJ* 2000;320:1530-3.
15. Weston A.,Fitzgerald P. Discrete choice experiment to derive willingness to pay for methyl aminolevulinate photodynamic therapy versus simple excision surgery in basal cell carcinoma. *Pharmacoeconomics* 2004;22:1195-208.
16. Phillips KA, Maddala T, Johnson FR. Measuring preferences for health care interventions using conjoint analysis: an application to HIV testing. *Health Services Research* 2002;37:1681-705.
17. Beusterien KM, Dziekan K, Flood E, Harding G, Jordan JC. (2005) "Understanding patient preferences for HIV medications using adaptive conjoint analysis: feasibility assessment", *Value Health*, 8(4):453-61.
18. Coast J, Salisbury C, de Berker D, Noble A, Horrocks S, Peters TJ et al. Preferences for aspects of a dermatology consultation. *BR J Dermatol* 2006;155:387-92.
19. King MT, Hall J, Lanscar E, Fiebig D, Hossain I, Louviere J et al. Patient preferences for managing asthma: results from a discrete choice experiment. *Health Econ* 2006;16:703-17.
20. Peacock S, Apicella C, Andrews L, Tucker K, Bankier A, Daly MB et al. A discrete choice experiment of preferences for genetic counselling among Jewish women seeking cancer genetics services. *B J Cancer* 2006;95:1448-53.
21. Roux L, Ubach C, Donaldson C, Ryan M. Valuing the benefits of weight loss programs: An application of the discrete choice experiment. *Obesity Research* 2004;12:1342-51.
22. Hauber AB, Johnson FR, Sauriol L, Lescrauwaet B. Risking health to avoid injections: preferences of Canadians with type 2 diabetes. *Diabetes Care* 2005;28(9):2243-5.
23. Johnson FR, Manjunath R, Mansfield CA, Clayton LJ, Hoerger TJ, Zhang P. High-risk individuals' willingness to pay for diabetes risk-reduction programs. *Diabetes Care* 2006;29(6):1351-56.
24. Marshall DA, Johnson FR, Phillips KA, Marshall JK, Thabane L, Kulin NA. (2007) "Measuring patient preferences for colorectal cancer screening using a choice-format survey", *Value Health*, 10(5):415-30.
25. Hauber AB, Johnson FR, Mohamed AF, Leibman C, Arrighi HM, Grundman M. Older Americans' risk-benefit preferences for modifying the course of Alzheimer's disease. *Alzheimer Dis Assoc Disord* 2009;23(1):23-32.
26. Drummond MF, O'Brien BJ, Stoddart GL, et al. *Methods for the economic of health care programmes*. 2nd ed. New York: Oxford University Press, 1997
27. Huber J, Zwerina K. The importance of utility balance in efficient choice designs. *Journal of Marketing Research* 1996;33(3):307-317.
28. Louviere JJ, Hensher DA, Swait JD. *Stated Choice Methods: Analysis and Applications*. New York: Cambridge University Press. 2000.
29. Hedayat AD, Sloane N, Stufken J. *Orthogonal Arrays: Theory and Applications*. New York: Springer-Verlag, 1999.
30. Han G, Shapiro S. A catalog and computer program for the design and analysis of orthogonal symmetric and asymmetric fractional factorial designs. Schenectady/NY. Report No. 66-C-165,GE Corp., 1966.

31. Carlsson F, Martinsson P. Design techniques for stated-preference methods in health economics. *Health Economics* 2003;12(4):281-294.
32. Lloyd AJ, McIntosh E, Williams AE, Kaptein A, Rabe KF. How does patient's quality of life guide their preferences regarding aspects of asthma therapy? *Patient* 2008; 1: 309-316
33. Orme BK. Getting Started with Conjoint Analysis: Strategies for Product Design and Pricing Research. Madison Wisconsin: Research Publishers LLC. 2006.
34. Oppe M, Szende A, de Charro F. Comparative review of Visual Analogue Scale Value Sets. in EQ-5D Value Sets: Inventory, Comparative Review and User Guide Vol. 2. EuroQoL Group Monographs (Szende A, Oppe M, Devlin N, Eds.): Netherlands: Springer. 2007
35. Lloyd AJ Threats to the estimation of benefit: Are preference elicitation methods accurate? *Health Econ* 2003; 12(5):393-402.
36. Kjaer T, Bech M, Gyrd-Hansen D, Hart-Hansen K. Ordering effect and price sensitivity in discrete choice experiments: need we worry? *Health Econ.* 2006; 15:1217-28.
37. Johnson FR, Hauber AB, Osoba D, Hsu M, Coombs J, Copley-Merriman C. Are chemotherapy patients' HRQoL importance weights consistent with linear scoring rules? A stated-choice approach. *Qual Life Res*2006;15:285-98.
38. Lloyd A, D Penson, S Dewilde, L Kleinman Eliciting patient preferences for hormonal therapy options in the treatment of metastatic prostate cancer *Prostate Cancer and Prostatic Diseases* 2007
39. Hauber AB, Mohamed AF, Watson ME, Johnson FR, Hernandez JE. Benefits, risk, and uncertainty: preferences of antiretroviral-naïve African Americans for HIV treatments. *AIDS Patient Care STDS* 2009;23(1):29-34.
40. McColl,E, Fayers,PM. Proxy assessments and context effects, in: *Assessing Quality of Life in Clinical Trials: Methods and Practice* 2nd edn (ed Fayers,P. M.;Hays,R. D.), Oxford University Press, Oxford, 2005 pp131 - 146
41. Wild D, Grove A, Martin M, Eremenco S, McElroy S, Verjee-Lorenz A, Erikson P. Principles of Good Practice for the Translation and Cultural Adaptation Process for Patient-Reported Outcomes (PRO) Measures: Report of the ISPOR Task Force for Translation and Cultural Adaptation. *Value Health* 2005; 8(2): 94-104.
42. Kleinman L, McIntosh E, Ryan M, Schmier J, Crawley J, Locke GR 3rd, De Lissovoy G. Willingness to pay for complete symptom relief of gastroesophageal reflux disease. *Archives of Internal Medicine* 2002 Jun 24;162(12):1361-6.
43. Johnson FR, Özdemir S, Mansfield CA, Hass S, Miller DW, Siegel CA, Sands BE. Crohn's disease patients' benefit-risk preferences: serious adverse event risks versus treatment efficacy. *Gastroenterology* 2007;133(3):769-79.
44. Johnson FR, Hauber AB, Özdemir S. Using conjoint analysis to estimate healthy-year equivalents for acute conditions: an application to vasomotor symptoms. *Value Health* 2009;12(1): 146-152
45. Marshall DA, Johnson FR, Phillips KA, Marshall JK, Thabane L, Kulin NA. Measuring patient preferences for colorectal cancer screening using a choice-format survey. *Value Health* 2007;10(5):415-30.