Emily Lancsar, PhD, Associate Professor, Centre for Health Economics, Faculty of Business and Economics, Monash University, Melbourne, Victoria, Australia

Deborah Marshall, PhD, MHSA, Canada Research Chair, Health Services and Systems Research; Associate Professor, Department of Community Health Sciences, Faculty of Medicine, University of Calgary; Alberta Bone and Joint Health Institute, Calgary, AB, Canada

Axel Mühlbacher, PhD, Harkness Fellow in Health Care Policy & Practice, Duke Clinical Research Institute/ Fuqua School of Business, Duke University, Durham, NC, USA and Professor, HS Neubrandenburg, Neubrandenburg, Germany

Dean Regier PhD, Assistant Member, Fred Hutchinson Cancer Research Center & Assistant Professor, Pharmaceutical Outcomes Research and Policy Program, University of Washington, Seattle, Washington, USA

Vik Kilambi, BA, Research Associate, RTI Health Solutions, RTI International, Research Triangle Park, NC, USA
This taskforce on experimental design follows the good research practice guidelines by the ISPOR Good Research Practices for the Application of Conjoint Analysis in Health Task Force.

Table 1 – A Checklist for Conjoint Analysis Applications in Health and Medicine

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Was a well-defined research question stated and is conjoint analysis an appropriate method for answering it?</td>
<td>6. Was the data collection instrument designed appropriately?</td>
<td></td>
</tr>
<tr>
<td>1.1. Was a well-defined research question or testable hypothesis articulated?</td>
<td>6.1. Were appropriate information about respondents collected (such as socio-demographic, attitudinal, health history/status, and treatment experience)?</td>
<td></td>
</tr>
<tr>
<td>1.2. Was the study perspective described and the study placed in any particular decision-making or policy context described?</td>
<td>6.2. Were the attributes adequately described and two necessary contextual information provided?</td>
<td></td>
</tr>
<tr>
<td>1.3. What is the justification for using conjoint analysis to answer the research question?</td>
<td>6.3. Was the level of burden of the data collection instrument appropriate? Were respondents informed, encouraged, and motivated?</td>
<td></td>
</tr>
<tr>
<td>2. Were the attributes and attribute levels and supported by evidence?</td>
<td>7. Was the data collection plan appropriate?</td>
<td></td>
</tr>
<tr>
<td>2.1. Were all important and relevant attributes identified (that is, supported by literature reviews, focus groups, or other scientific method)?</td>
<td>7.1. Was the sampling strategy justified (for example, sample size, stratification, and recruitment)?</td>
<td></td>
</tr>
<tr>
<td>2.2. Were the scales of included attributes justified and consistent with theory?</td>
<td>7.2. Were the modes of administration justified and appropriate (for example, face-to-face, pen-and-paper, web-based)?</td>
<td></td>
</tr>
<tr>
<td>2.3. Were the range and number of levels for each included attribute justified?</td>
<td>7.3. Were human-subjects considerations addressed (for example, recruitment, information, and consent, compensation)?</td>
<td></td>
</tr>
<tr>
<td>3. Was the construction of the conjoint tasks appropriate?</td>
<td>8. Were statistical analyses and model estimation appropriate?</td>
<td></td>
</tr>
<tr>
<td>3.1. Was the number of attributes in each conjoint task justified (full profile or partial profile)?</td>
<td>8.1. Were respondent characteristics examined and tested?</td>
<td></td>
</tr>
<tr>
<td>3.2. Was the number of scenarios in each conjoint task justified?</td>
<td>8.2. Was the quality of the responses examined (for example, rationality, validity, reliability)?</td>
<td></td>
</tr>
<tr>
<td>3.3. Were the number of conjoint tasks included in the data collection instrument appropriate?</td>
<td>8.3. Was multi-response analysis conducted appropriately? Were issues of clustering and sub-groups handled appropriately?</td>
<td></td>
</tr>
<tr>
<td>4. Was the choice of experimental design justified and evaluated?</td>
<td>9. Were the results and conclusions valid?</td>
<td></td>
</tr>
<tr>
<td>4.1. Was the choice of experimental design justified? Were alternative experimental designs considered?</td>
<td>9.1. Did results reflect testable hypotheses and account for statistical uncertainty?</td>
<td></td>
</tr>
<tr>
<td>4.2. Were the properties of the experimental design evaluated?</td>
<td>9.2. Were conclusions supported by the evidence and compared to existing findings in the literature?</td>
<td></td>
</tr>
<tr>
<td>4.3. Was (should) an opt-out or a status-quo alternative (be) included?</td>
<td>9.3. Were study limitations and generalizability adequately discussed?</td>
<td></td>
</tr>
<tr>
<td>5. Were preferences elicited credibly?</td>
<td>10. Were the study presented well and completely?</td>
<td></td>
</tr>
<tr>
<td>5.1. Were the conjoint tasks sufficiently motivated and explained?</td>
<td>10.1. Was the study importance and research context adequately motivated?</td>
<td></td>
</tr>
<tr>
<td>5.2. Was an appropriate elicitation format (the choice is, rating, ranking, or choice) used? Also, did (should) the elicitation format allow for indifference?</td>
<td>10.2. Were the study methods explained and the data collection instrument adequately described and/or illustrated?</td>
<td></td>
</tr>
<tr>
<td>5.3. In addition to preference elicitation, did the conjoint tasks include other questions (e.g., strength of preference, confidence in response, other method)?</td>
<td>10.3. Were the implications of the study stated and understandable to a wide audience?</td>
<td></td>
</tr>
</tbody>
</table>
While conjoint analysis & discrete-choice experiments have intuitive appeal, it is easy to underestimate the complexity of experimental design.

- Consider 3x3x3x3x3x3 attribute/level space
  - 729 profiles
  - 531,441 pairs or 387,420,489 triplets

- If one wanted to have a design that consisted of 20 choice tasks you could generate:
  - $3.1 \times 10^{114}$ survey versions with pairs
  - $5.8 \times 10^{171}$ survey versions with triplets

---

**Emily Lancsar, PhD**  
Associate Professor  
Centre for Health Economics  
Faculty of Business and Economics  
Monash University  
Melbourne, Victoria, Australia

- Establish good research practices for experimental design for DCE applications in health
**Rationale**

- DCE routinely used in health & health care
- Experimental design crucial component of any DCE
- Perhaps the part of undertaking a DCE about which researchers are often most unsure
- Rapidly moving literature, little consensus

**Experimental Design**

- Combination of attributes & levels to create hypothetical options & placement of options into choice sets
- Respondents asked to choose between the options in each choice set
**Example Choice Set**

<table>
<thead>
<tr>
<th>Features</th>
<th>Treatment Plan A</th>
<th>Treatment Plan B</th>
<th>Current Treatment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Diabetes medications:</td>
<td>90-120 mg/dL</td>
<td>70-170 mg/dL</td>
<td>&gt; 170 mg/dL</td>
</tr>
<tr>
<td>Fasting glucose:</td>
<td>&lt; 7%</td>
<td>7-8%</td>
<td>&gt; 8%</td>
</tr>
<tr>
<td>HbA1c</td>
<td>3-6</td>
<td>&lt; 1</td>
<td>1-2</td>
</tr>
<tr>
<td>Hypos/month</td>
<td>90-120 mg/dL</td>
<td>&gt; 170 mg/dL</td>
<td></td>
</tr>
<tr>
<td>Out-of-pocket cost/month</td>
<td>$100</td>
<td>$50</td>
<td>$150</td>
</tr>
</tbody>
</table>

Which treatment would you choose if these were the only options available?

- A
- B
- Current

**ED & Data Analysis Directly Linked**

- $U_{ij} = \beta x_{ij} + \epsilon_{ij}$
- Experimental design provides data for RHS variables (ignoring socio-demos for now)
- Respondent choice => dependent variable
- So, when using a DCE we have most of our data before going to field (unlike RP data)
**Importance of Experimental Design**

- Determines the statistical properties of the RHS variables (data matrix)
- Has key implications for composition & properties of regression models estimated from DCEs
- Onus on researchers to ensure high quality ED, i.e. high quality data
- It matters!

**Design Objectives**

- Identification
- Statistical Efficiency
- Response Efficiency
Which effects ($\beta_X$) can be independently estimated from the experimental design

E.g. main effects for each attribute, interactions between attributes etc.

- Precision with which the effects of interest ($\beta_X$) can be estimated
- Lower confidence intervals around effects of interest
- Can increase precision by increasing sample size
- Matters more for small samples
- D error & D optimal
**Response Efficiency**

- Linked to task difficulty
- Level of attention, fatigue, boredom
- Implausible combinations
- Overlap
- Possible tradeoff statistical & response efficiency

**Key Design Issues in a Health Context**

- Implausible attribute level combinations
- Difficulty separating some attributes – e.g. pain and duration
- Cognitive ability of some respondent groups
- Generic v labelled alternatives
- Not all unique to health
Approaches

- Manual construction
- Design catalogues – websites
- SAS
- Sawtooth
- Street & Burgess
- Sandor & Wedel
- Rose & Bleimer

Comparison

- Approaches differ in relation to:
  - Model specification
  - Level of importance placed on knowing the optimum
  - Level of importance placed on orthogonality
  - Priors on beta coefficients
  - Ease of including user defined constraints
  - Cost – many design options are free, others must be paid for
  - Software availability
  - Ease of use
So the preferred design approach is....

**Preferred Approach**

- Which approach is “best” depends on each study’s specific design objectives and preferences of the researcher
- We have provided information to allow researchers to make (more) informed decisions re choice of experimental design approach
- Rapidly evolving area
Applications of DCE methods require developing experimental designs. Many health applications involve specific considerations:

- 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
- Blocking
Choice data are collected using health profiles based on hypothetical alternatives.

- Some possible attribute-level combinations could be implausible or inconsistent with logical expectation.

Example: a design with two attributes
- *activities of daily living* (no restrictions vs. some restrictions) and
- *symptoms* (mild vs. moderate vs. severe).

- Conjoint task that asks a respondent to evaluate a treatment alternative that combines no restrictions with severe symptoms would result in an implausible scenario or outcome.

Illogical combinations could increase:
- the potential for hypothetical bias
- unobserved, heterogeneous interpretations by respondents
- lower response efficiency

- Some design approaches allow researchers to specify combinations that should not appear in the design, while other approaches do not.
Associated list of relevant attributes may include scenarios where interactions are likely among different attributes.

Symptom severity and duration often are, in effect, a single compound attribute with two dimensions.

- Respondents cannot evaluate outcomes where severity and duration are treated as separate attributes.

Example:

- 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.

The balance between acceptable response efficiency and statistical efficiency may have to favor simpler designs.

- Yield less statistical information for a given sample size
**Cognitive Limitations of Specific Groups of Respondents**

- Choice questions are cognitively challenging.
  - Statistically efficient designs may be beyond the reach of certain respondents
    - e.g., respondents with a condition that involves cognitive deficits such as Alzheimer’s disease, schizophrenia, or other neurological conditions

---

**Cognitive Limitations of Specific Groups of Respondents**

- The balance between acceptable response efficiency and statistical efficiency may have to favor simpler designs
  - yield less statistical information for a given sample size
The majority of DCE studies have used experimental designs with generic choice alternatives (e.g., Medicine A, Medicine B).

Choice alternatives also can be given labels, where the label has some meaning (e.g., nurse practitioner, general practitioner).

LMA designs
- incorporate labeled alternatives
- allow for the independent estimation of alternative-specific attribute effects
- can be created by following standard approaches used to created generic designs, but alternative-specific attributes are treated as separate design columns
- simultaneously create both the alternatives and the choice questions
Example 1: 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.

Example 2: 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.

Many software packages can accommodate them via internal options or through the ability to specify user-defined constraints.

- Discrete Choice Experiments (Burgess L, University of Technology, Sydney)
- Ngene (Rose JM, Bliemer MCJ)
- Sawtooth Software (Sequim: Sawtooth Software)
- Gauss 8.0 (Black Diamond: Aptech Systems Inc)
An experimental design that is constructed prior to fielding might contain more choice questions than the researcher wishes to ask to each respondent.

- Blocking of the experimental design should be carefully considered.

Blocks = Partitions of the choice questions (usually equally sized) 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.

Example: An experimental design with 24 choice questions can be partitioned randomly into 2 blocks with only 12 choice questions each.

**Advantage:**
Blocking promotes response efficiency by reducing the necessary cognitive effort for each respondent.

**Disadvantage:**
Desirable statistical properties of the experimental design (e.g., no correlations among attribute levels) 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.

Alternatively, the researcher may remove the need for blocking altogether by randomly sampling a number of choice questions in the experimental design for each respondent.

F. Reed Johnson PhD
Chair, ISPOR Conjoint Analysis Experimental Design Task Force Chair and Distinguished Fellow and Principal Economist, Health Preference Assessment Group, RTI Health Solutions, RTI International, Research Triangle Park, NC, USA
**Task Force Strategy**

- Readers with some experience with DCE methods
- Non-technical summary of basic concepts
- Focus on decisions required for empirical research
- Avoid evaluating or endorsing particular software
  - Capabilities
  - Ease of use
  - Accessibility

**Efficiency measures**

- Confusion about D-efficiency measures
- Report now makes a distinction between relative and absolute D-efficiency
  - Absolute D-efficiency depends on variable coding, model specification, and priors
  - Relative D-efficiency compares multiple designs within a class
  - Researchers need to use each consistently and appropriately
Too much initial focus on pharmaceutical evaluations

Report now acknowledges differences in applications involving outcomes, technologies, process, and policy

Different significance of:
- Cognitive limitations of particular respondent groups
- Labeled and constant alternatives for interventions and technologies

Questions on exclusive focus on DCE

Now acknowledges basic principles apply more widely
- Best-worst scaling
- Rating and ranking
- Combining revealed- and stated-preference data
Detail On Particular Implementations

- Objections to neutral, high-level guidance on some approaches
- Accepted refinement of draft text, but not significant elaboration of software capabilities

Comments I

- DCE versus conjoint
- Level of sophistication of conceptual section
- Format and content of comparison tables
  - Deleted
- Full choice or full factorial?
  - Full choice. Draw only from set of feasible choices
### Comments II

- Preference weights or choice simulation?
  - Preference weights

- Design implications of linear and nonlinear models
  - Clarified importance of designing for nonlinearity

- Estimating models different than assumed for the design changes efficiency
  - Clarified importance of designing for flexible modeling options

### Comments III

- Differences among choices, judgments, and preferences
  - Observe choices or judgments. Preferences inferred

- “respondents”, not “patients”
  - Yes. What about “subjects”? Acknowledged

- Pivot designs, role of status quo or reference condition
  - Acknowledged
**Task Force members**: Emily Lancsar, Deborah Marshall, Vikram Kilambi, John Bridges, Axel Mühlbacher, Dean Regier, Brian Bresnahan, Barbara Kanninen

**Outside reviewers**: Brett Hauber, Ben Craig, Bryan Orme, Warren Kuhfeld, John Rose, Christine Hutton, Deborah Street, Zsolt Sándor, Leonie Burgess, Michel Wedel, Sheryl Szienbach, Jo Mauskopf

**and particularly** Elizabeth Molsen