



**10th Meeting “International Academy of Health Preference Research” Basel, Switzerland, 13–14 July, 2019**

Published online: 18 June 2019

**ABSTRACTS**

**10th Meeting  
International Academy of  
Health Preference Research  
Basel, Switzerland  
13–14 July, 2019**



## 10TH MEETING OF THE INTERNATIONAL ACADEMY OF HEALTH PREFERENCE RESEARCH

**Axel C. Mühlbacher**

Established on 15 April 2014, the International Academy of Health Preference Research (IAHPR) is a member-driven, inter-generational organization that promotes educational activities and research with respect to health and health-related preferences. Our aim is to improve decisions about health and healthcare throughout the world by developing, promoting, and supporting health preference research with the widest possible applicability.

The 10th Meeting of the International Academy of Health Preference Research will be held on Saturday and Sunday, 13–14 July 2019 at the Volkshaus in Basel, Switzerland. Chaired by Esther W. de Bekker-Grob and Jennifer A. Whitty and hosted by Axel C. Mühlbacher, its activities include a workshop, a symposium, a networking dinner, and a scientific meeting.

On 13 July 2019, the Academy and Patient Preferences in Benefit and Risk Assessments during the Treatment Life Cycle (PREFER) will host a joint morning workshop on “Good research practices for health preference studies,” led by Axel C. Mühlbacher. This workshop will describe the basic on how to conduct a health preference study focusing on trade-offs between risks and benefits. IAHPR members will provide examples of challenges faced during the assessment of patient preferences in health care decision making. The workshop material will build directly from the textbook under development by IAHPR members, incorporating the experiences of scientists working with PREFER.

After lunch, the Academy and PREFER will host a joint afternoon symposium on “Patient preferences in medical treatment lifecycle.” This topic is of high relevance for the objectives of both the Academy and PREFER. After the presentations by invited speakers, the panel will discuss critical topics defined in advance by the co-chairs, followed by a question and answer session. The symposium discussion will be summarized for publication in *The Patient*, an official journal of the IAHPR. After the symposium, the Academy and PREFER will host a joint networking dinner.

Starting at 8:00 on Sunday, 14 July 2019, the Academy will host the scientific meeting, including twelve podium presentations, lunch (with a poster session), and a business session. Twenty-seven abstracts were submitted for this meeting. Each was blinded then rated by 38 of the 44 tenured members of the Academy. The twelve abstracts with the highest rating were invited for podium presentation and are listed chronologically.

### Disclaimer

IAHPR in general requests that a high standard of science is followed concerning publications and presentations at all its workshops, symposia and meetings. However, IAHPR as a whole or its Foundation, or its members, do not take any responsibility for the completeness or correctness of data or references given by authors in publications and presentations at IAHPR events.

It is not within the remit of IAHPR or its Foundation, in particular, to seek clarification or detailed information from authors about data in submitted abstracts. Moreover, it is not within the scope of IAHPR and its committees to monitor compliance with any legal obligations, e.g., reporting requirements or regulatory actions.

## Beating the Benchmarks: Using Patient Preferences to Increase the Probability of Development Success

B. S. Levitan<sup>1</sup>, E. G. Katz<sup>1</sup>, R. L. DiSantostefano<sup>2</sup>, J. C. Yang<sup>3</sup>,  
A. O. Fairchild<sup>3</sup>, S. D. Reed<sup>3</sup>, F. R. Johnson<sup>3</sup>

<sup>1</sup>*Epidemiology, Janssen R&D, Titusville, NJ, USA;*

<sup>2</sup>*Epidemiology, Janssen R&D, Raritan, NJ, USA;*

<sup>3</sup>*Duke Clinical Research Institute, Duke University, USA*

**Background:** Drugs in development have notoriously low benchmark probabilities to reach the market. A key step in navigating these low probabilities is defining strategic requirements for development success. An industry strategy document, the target product profile (TPP), specifies minimum requirements for efficacy, safety, tolerability, formulation, dosing and other drug properties. If the TPP goals are met, development proceeds. If not, the compound strategy is reconsidered, forecasts are revised, and development may be halted.

**Methods:** While the concept of alternatives forms of success is intuitive, TPPs generally specify just one or a few options. The challenge is having a defensible means to specify equally valued alternatives. We used findings from two preference studies to show how assessing maximum-acceptable risk (MAR) for a range of benefits can generate a large family of preferentially equivalent alternatives: (1) a preference study that assessed the MAR of sudden death or disabling stroke in exchange for delaying the onset of Alzheimer's disease. (2) A preference study in treatment-resistant depression (TRD) that estimated the MAR of permanent memory/cognitive and bladder problems for improvements in depression.

**Results:** In the Alzheimer's study, for 1-year delayed onset, participants would accept 5% chance of disabling stroke. For 2 years delay, 11%. For 3 years, 17%. In the TRD study, we calculated joint probabilities of memory/cognitive problems and bladder problems that would be acceptable for different levels of benefit. For improvement from moderate to mild depression, patients would accept joint (memory/cognitive, bladder) MARs of (1.9%, 0), (1%, 1.3%), (0, 2.7%) and many other combinations. For improvement from severe to mild depression, the joint MARs are higher and include (5.1%, 0), (3%, 3%) and (0, > 5%).

**Conclusions:** Preference studies can give a large family of TPP trade-offs equally valued by patients and with similar market share. These define alternative paths for development success that can "beat the benchmarks" and increase the probability of development success.

## Valuation space models for the analysis of choice experiments: an example in exome sequencing

D. A. Marshall<sup>1</sup>, K. V. MacDonald<sup>1</sup>, S. Heidenreich<sup>2</sup>,  
K. M. Boycott<sup>3</sup>

<sup>1</sup>*University of Calgary, Calgary, Alberta, Canada;*

<sup>2</sup>*Health Economics Research Unit, University of Aberdeen, Aberdeen, Scotland; Evidera, Inc., London, UK;*

<sup>3</sup>*Children's Hospital of Eastern Ontario Research Institute, University of Ottawa, Ottawa, Ontario, Canada*

**Background:** Mixed logit models for the analysis of health care choices usually estimate random marginal utilities. Marginal rates of substitutions (MRSs) are subsequently obtained as the ratio of two coefficients. To ensure that obtained distributions of MRSs have finite moments, the

distribution of the numéraire needs to be fixed or bounded. However, resulting ratio distributions can be highly skewed, behaviourally implausible or difficult to interpret. Previous research suggests overcoming these limitations by directly estimating distributions of MRSs. Using a discrete choice experiment (DCE) estimating the added value of exome sequencing (ES) over standard diagnostic tests for rare diseases, we illustrate the usefulness of such valuation space models.

**Methods:** We administered a DCE with six attributes (diagnostic test, chance of diagnosis, negative impact of diagnosis, positive impact of diagnosis, out of pocket test cost and time to diagnosis) to parents of children with rare diseases. Valuation-space models were used to obtain three MRSs: willingness to pay, willingness to wait for test results and minimum acceptable chance of a diagnosis.

**Results:** 89% of 319 respondents reported their child had genetic testing, 66% received a diagnosis and 26% reported that their child had been offered ES. For most attributes, preferences varied significantly between respondents. The valuation-space model results estimated that parents would be willing to pay CAD\$6590 (SD: \$5050), wait 5.2 years (SD 3.98 years) to obtain a diagnostic test result, or accept a reduction of 3.1% (SD 2.44%) in the chance of receiving a diagnosis for ES testing compared to operative procedures.

**Conclusions:** While random marginal utilities can account for unobservable heterogeneity in preferences, distributions or MRSs can be highly skewed and may require unreasonable assumptions to ensure model identification. Valuation-space models can meaningfully address this problem by directly estimating the distributions of MRSs.

## Preferences in Precision Medicine: Biomarker-Based Treatment to Delay Type-1 Diabetes

R. DiSantostefano<sup>1</sup>, J. Sutphin<sup>2</sup>, K. Gallaher<sup>2</sup>, C. Mansfield<sup>2</sup>

<sup>1</sup>*Janssen R&D, LLC, Titusville, NJ, USA;*

<sup>2</sup>*RTI Health Solutions, RTP, NC*

**Background:** Biomarker screening and associated treatment decisions to prevent or delay disease involve layers of uncertainty and complexity, and they are increasingly utilized in personalized and preventive medicine. We evaluated parent preferences for hypothetical treatments that delay the onset of T1D insulin dependence in children to inform medicines development.

**Methods:** A discrete choice experiment survey using an online research panel assessed the preferences of US parents told to assume one of their children (< 18 years) would become insulin dependent with T1D within 2 years based on a biomarker test. The online web-based panel ( $n = 1501$ ) included parents with ( $n = 600$ ) and without ( $n = 901$ ) a child with T1D. Respondents were offered a series of eight choices between two hypothetical treatments that would delay T1D or an opt out (monitoring only). Treatments were defined by six attributes with varying levels of benefits and harms. Random Parameter Logit (RPL) modeling was used to assess preferences, stratified by already having/not having a child with T1D. Latent class analysis (LCA) was used to explore heterogeneity.

**Results:** Most parents chose a treatment (2% always chose the opt out). LCA results yielded 5 classes where parents focused mostly on (1) delaying T1D insulin dependence, (2) reducing long-term risk of T1D complications, (3) avoiding serious infection, (4) monitoring only (opt out), (5) and a disordered class (~ 20%) that may have based their decision on other properties, misunderstood, and/or were task non-

attendant. Class membership was related to differences in patient characteristics, insurance status, and performance on comprehension questions. **Conclusions:** This study identified five distinct groups whose preferences can inform development decisions for future treatments to delay T1D. The growth of precision medicine requires understanding preferences in a more complex and uncertain decision context, which may require advancements in preference methods.

## Can Healthcare Choice be Predicted Using Stated Preference Data?

E. W. de Bekker-Grob<sup>1</sup>, B. Donkers<sup>1</sup>, M. C. J. Bliemer<sup>2</sup>, J. Veldwijk<sup>1</sup>, J. D. Swait<sup>1</sup>

<sup>1</sup>Erasmus Choice Modelling Centre, Erasmus University Rotterdam; <sup>2</sup>Business School, University of Sydney

**Background:** The lack of evidence about the external validity of discrete choice experiments (DCEs) is one of the barriers that inhibits greater use of DCEs in healthcare decision-making. This study examines external validity of DCE-derived preferences, unravel its determinants, and provide evidence whether healthcare choice is predictable.

**Methods:** We focused on the field of influenza vaccination and used a six-step approach: (1) a literature study, (2) expert interviews, (3) focus groups, (4) a survey including a DCE, (5) field data, and (6) in-depth interviews with respondents who showed discordance between stated preferences and actual healthcare utilization as a mean of diagnosing model mis-specification. Respondents without missing values in the survey and the actual healthcare utilization ( $377/499 = 76\%$ ) were included in the final analyses. Random-utility-maximization and random-regret-minimization choice processes were used to analyze the DCE data, whereas the in-depth interviews combined five scientific theories to explain discordance.

**Results:** When models took into account both scale and preference heterogeneity, real-world choices to opt for influenza vaccination were correctly predicted by DCE at an aggregate level, and almost 90% of choices were correctly predicted at an individual level. There was 13% ( $49/377$ ) discordance between stated preferences and actual healthcare utilization. In-depth interviews showed that several dimensions played a role in clarifying this discordance: attitude, social support, action of planning, barriers, and intention.

**Conclusions:** Evidence was found, at least in this particular study, that DCE yields accurate predictions of real-world behavior if at least scale and preference heterogeneity are taken into account. Analysis of discordant subjects showed that we can even do better. The DCE measures an important part of preferences by focusing on attribute tradeoffs that people make in their decision to participate in a healthcare intervention. Inhibitors may be among these attributes, but it is more likely that inhibitors have to do with exogenous factors like goals, religion, phobias, and social norms. Conducting upfront work on constraints/inhibitors of the focal behavior, not just what promotes the behavior, might further improve predictive ability.

## Number of Halton Draws Required for Valid Random Parameter Estimation with Discrete Choice Data

A. Ellis<sup>1</sup>, E. de Bekker-Grob<sup>2</sup>, K. Howard<sup>3</sup>, K. Thomas<sup>4</sup>, E. Lancsar<sup>5</sup>, M. Ryan<sup>6</sup>, J. Rose<sup>7</sup>

<sup>1</sup>Department of Social Work, North Carolina State University, Raleigh, USA;

<sup>2</sup>Erasmus School of Health Policy and Management, Erasmus University Rotterdam, Netherlands;

<sup>3</sup>School of Public Health, University of Sydney, Australia;

<sup>4</sup>UNC Eshelman School of Pharmacy, University of North Carolina at Chapel Hill, USA;

<sup>5</sup>Department of Health Services Research and Policy, Australian National University;

<sup>6</sup>Health Economics Research Unit, University of Aberdeen, UK;

<sup>7</sup>Business Intelligence and Data Analytics Research Centre, University of Technology Sydney, Australia

**Background:** Mixed-logit models of discrete choice experiment (DCE) data often simulate random parameters with Halton draws. The model assumes uncorrelated random parameters with certain (often normal) distributions. Using too few draws may violate these assumptions, biasing estimates and standard errors, but guidance about number of draws is lacking. Systematic review data show that number of draws is rarely reported, highly variable, and unrelated to number of random parameters. We developed guidance about the number of Halton draws to use in these models.

**Methods:** In R, we simulated random parameters using 50 Halton sequences with 50 to 10,000 draws. We (1) plotted normality test results, (2) plotted correlations among parameters, (3) assessed bias and relative efficiency in real data, using models with 5, 10, and 15 random parameters and 250 to 20,000 draws, and (4) evaluated current practice by overlaying plots with data on modeling practices from 40 DCEs.

**Results:** Univariate normality: With 500 draws and 10 random parameters, or 1000 and 12, one random parameter departed from normality. With 500 draws and 17 random parameters, or 1000 and 22, half departed from normality. Multivariate normality: With  $\geq 7$  random parameters, the Henze–Zirkler p-value decreased. With 11, keeping  $p > 0.05$  required 4000 draws. Based on actual modeling practices, 16/40 recently published DCEs (40%) likely used insufficient draws for multivariate normality. Correlations among random parameters: Keeping correlations  $< 0.2$  required 250 draws when there were 10–15 random parameters and 1000 draws when there were 22 random parameters. Based on actual modeling practices, 5/40 recent DCEs (13%) likely had correlations  $> 0.1$  and 2/40 (5%) likely had correlations  $> 0.2$ , violating model assumptions. Real data: Models with more random parameters and fewer draws yielded bias and incorrect standard errors. With 15 random parameters, all estimates were unstable.

**Conclusions:** Stable mixed-logit estimation requires  $< 10$  random parameters and  $> 1000$  draws. Among 40 recent DCEs, 14 (35%) met both conditions. Future studies should develop specific guidelines and explore alternative methods. Meanwhile, number of draws should increase with number of random parameters, exceed customary levels, and be reported. Analysts should use sufficient draws for all analyses, then use more draws to verify final results. Insufficient draws may bias estimates, standard errors, and healthcare decisions.

## LC vs. SALC: Choosing Between Latent Class Models of Preference Heterogeneity

S. Karim<sup>1</sup>, B. M. Craig<sup>1</sup>, S. Poteet<sup>1</sup>

<sup>1</sup>University of South Florida

**Background:** In choice modeling, the existence of heterogeneity in structural preferences (i.e., trade-offs) and in variance (scale) (Groothuis-Oudshoorn et al. 2018) creates a dilemma for preference researchers: latent class (LC) or scale-adjusted latent class (SALC)? LC models create classes mixing both forms simultaneously, and SALC models separate them into two class types (trade-off and scale). The objective of this paper is to examine the performance of the LC and SALC models using a case example, the demand for health insurance plans.

**Methods:** The analysis included five sets of variables: paired comparison responses, plan attributes, respondent characteristics, current plan characteristics, and behavioral characteristics. The LC model identified its classes using all three characteristics, and the SALC model identified the trade-off classes using respondent and plan characteristics and the scale classes using respondent and behavioral characteristics. All models were estimated using Latent Gold (Magidson 2019). The optimal number of classes was set using the Bayesian information criterion (BIC).

**Results:** Analyzing the different LC and SALC models, the dilemma is between the LC with 3 classes (BIC 58136) and the SALC with 2 trade-offs/2 scales (58043). The two of the LC classes look similar, except one of has mis-ordered levels and smaller parameters. Respondents with less education, finished in less than 10 min, were more likely to belong to the class with the mis-ordered parameters. The SALC results clearly showed the distinction of between the two trade-off classes and between the two scale classes. Lastly, we compared the LC and SALC classes and found that the second trade-off class of the SALC looks like the merger of the two LC classes, except without the mis-ordered, small parameters.

**Conclusions:** The study demonstrates a case where the SALC model greatly improved the interpretation of preference heterogeneity (both forms). Future studies may attempt to incorporate respondent education and survey duration into their SALC models.

### References:

Groothuis-Oudshoorn et al. Key issues and potential solutions for understanding healthcare preference heterogeneity free from patient-level scale confounds. *Patient*. 2018;11(5):466–6.  
Magidson J. Latent gold. Belmont: Statistical Innovations; 2019.

## Benefit–Risk or Risk–Benefit Trade-offs? Another Look at Attribute Ordering Effects in DCEs

S. Heidenreich<sup>1,2</sup>, A. Beyer<sup>3</sup>, B. Flamion<sup>4</sup>, M. Ross<sup>1</sup>, J. Seo<sup>1</sup>, K. Marsh<sup>1</sup>

<sup>1</sup>Evidera Inc, London;

<sup>2</sup>University of Aberdeen;

<sup>3</sup>Innovus Consulting, London;

<sup>4</sup>Idorsia Pharmaceuticals Ltd

**Background:** Discrete choice experiments (DCEs) are increasingly used for health care valuation. Policy makers (i.e. regulators and payers) have signaled their interest in exploring the use of patient preference data from DCEs in benefit-risk assessments. Using DCEs for policy making raises questions about the effect of design aspects on collected data. We use a

pilot DCE, which will be integrated into a Phase 3 trial evaluating a new insomnia treatment, to explore the effect of attribute ordering on data quality indicators and statistical error variance. Only few studies previously assessed the effect of attribute ordering in DCEs and none within a benefit-risk context.

**Methods:** Respondents (N = 200) were randomized between three attribute orderings: (1) random; (2) benefits presented before risks; and (3) risks presented before benefits. Respondents were asked to complete 12 choices between unlabeled treatments and were given an opt-out option. Data quality and validity assessments included a dominance test, a preference stability test, numeracy scores, health literacy scores, and choice certainty. The effect of attribute ordering on error variance was assessed in a random effects model with design specific constants and scale heterogeneity.

**Results:** While we found no significant difference in observable data quality and internal validity measures, attribute ordering had a significant effect on the error variance. This suggests that attribute ordering may affect how respondents completed or interpreted the DCE. The error variance decreased significantly with deterministic ordering, compared to random attribute presentation. Error variance increased with the variability of stated choice certainty, health literacy, and numeracy.

**Conclusions:** Future applications of DCE should explore the implication of presentation order during instrument development. Future methods work should assess the effect of attribute ordering on policy advice and on respondents' decision-making process. Funding This study was funded by Idorsia Pharmaceuticals Ltd.

## Preferences for Exercise and Nutrition Programs: A Menu Choice Stated Preference Task

E. Lancsar<sup>1</sup>, E. Huynh<sup>1</sup>, J. Swait<sup>2</sup>, J. Ride<sup>3</sup>

<sup>1</sup>ANU;

<sup>2</sup>Erasmus;

<sup>3</sup>University of Melbourne

**Background:** DCEs typically elicit a single choice from presented options. However, health programs/services often can or must be combined in bundles (e.g. bundling private health insurance; packaging of care coordination). We present an adaption to standard DCEs to allow for synergies between programs, to appropriately measure demand and improve external validity of the task. Our contribution is two-fold: (1) methodologically, we present a menu-based experiment to explore bundling in the context of nutrition and exercise programs; (2) econometrically, we analyse the menu-based data using an extension of the choice set generation model (GenL) proposed by Swait (2001) to account for the potential for individuals to engage in choice set formation.

**Methods:** In an online menu-based experiment, respondents were presented with three programs: a nutrition program, an exercise program and their current status quo. Respondents could choose: the nutrition program (N); the exercise program (E); both nutrition and exercise programs (C); or their status quo (S). Programs were described by cost, average weight loss, program duration and incentives, plus exercise and nutrition program-specific attributes. MNL and GenL models were compared.

**Results:** A nationally representative sample of 333 Australians completed the survey. Overall, the best GenL model performed better than the MNL ( $\chi^2 = 58.99$ , 5 df,  $p < 0.001$ ). The MNL incorrectly assumes 100% weighting on the full choice set {N, E, C, S}, which accounted for only 39% of the choice set probabilities on average across the sample in the GenL. Consideration of bundling nutrition and exercise programs jointly



accounted for 69% ( $p < 0.001$ ) of choice set probabilities on average across the sample.

**Conclusions:** We provide a template for adapting DCEs and their analysis to capture bundling options using the case study of exercise and nutrition, where programs are potentially complementary in achieving the desired goal of improving health.

### An Embarrassment of Riches: What Can You Do with 10,000 Observations?

F. R. Johnson<sup>1</sup>, J. M. Gonzalez<sup>1</sup>, J. C. Yang<sup>1</sup>, J. Weatherall<sup>2</sup>, S. Kymes<sup>2</sup>

<sup>1</sup>*Department of Population Health Sciences, Duke University;*  
<sup>2</sup>*Lundbeck*

**Background:** The value of health spending depends on the public's willingness to pay higher taxes or reduce non-health program expenditures. Heterogeneity in preferences for taxes and programs raises questions about how to identify policy-relevant health-expenditure values. Health-policy questions also may require larger samples than commonly found in the discrete-choice experiment (DCE) health literature to inform priority-setting decisions.

**Objective:** To apply latent-class analysis using a very large data set to account for a large number of location-specific preference correlates.

**Methods:** 10,000 US adults completed an online DCE survey. Respondents answered 5 3-alternative trade-off questions consisting of status quo and two budget alternatives. Each budget profile included a mental-health program plus 2 programs randomly selected for each respondent from 4 programs: food safety, disaster relief, unemployment, and motor-vehicle safety. Benefits were scaled proportional to state population sizes. Modeling included split-sample, conditional and random-parameters logit, and various latent-class specifications, including predetermined and unconditional class assignments, with and without random parameters, with and without scale adjustments, with and without covariates, and with and without attribute-covariate interactions.

**Results:** Aggregate, split-sample, and latent-class analysis with predetermined-classes by state size yielded highly significant, but disordered, effect-coded coefficients and implausible value estimates. Unconditional latent-class models explained the implausible aggregate estimates as the result of averaging highly heterogeneous group preferences. Plausible latent classes included groups who rejected taxation for any purpose (21% of the sample), who approved taxation for any purpose (14%), who had well-defined priorities among programs and were: highly sensitive to (24%), ignored (21%), or less sensitive to tax increases (20%). Only the latter group passed a scope test on tax levels.

**Conclusions:** A rare opportunity to analyze a very large DCE dataset offered numerous options for well-powered hypothesis tests but also presented challenges in how to interpret and aggregate dissimilar preferences to support decision making.

### What if 0 is Not Equal to 0? Inter-personal Utility Anchoring Using the Worst Fears

M. K. Jakubczyk<sup>1</sup>, D. Golicki<sup>2</sup>

<sup>1</sup>*SGH Warsaw School of Economics;*  
<sup>2</sup>*Medical University of Warsaw*

Our worst fears differ. Some people dread death while others are horrified of pain. Utilities can be rescaled within any individual but the inter-personal comparisons are questionable. Still, when compiling valuations by multiple respondents the utility of dead is assumed identical across individuals:  $u(\text{dead}) = 0$ . We motivate another approach: we assume the difference between the worst health state (as defined by EQ-5D-5L plus dead) and the best one (11111), i.e. the maximal possible improvement, is equal between individuals. Then the disutilities of dimensions/levels/dead are estimated in such range. The resulting population means are rescaled, so average  $u(\text{dead}) = 0$  for convenience. Our approach has intuitive properties. Say, one respondent thinks moving from dead to perfect health (11111, i.e.  $\text{dead} \rightarrow 11111$ ) for a year is worth twice as much as  $55555 \rightarrow 11111$ , and another respondent thinks the exact opposite. Intuitively, they collectively value the improvements as equal. However, in utility terms, we would write  $u(55555) = -1$  and  $0.5$ , respectively. Hence,  $u(55555) = -0.25$  on average, and  $55555 \rightarrow 11111$  delivers larger utility gain than  $\text{dead} \rightarrow 11111$ . In comparison, our approach yields  $u(\text{dead}) = u(55555) = 0$ . We test our approach using Polish EQ-5D-5L data (TTO only, 1252 individuals, 11,480 observations). Being dead was strictly the worst fear in 30% of individuals, and for 63% there was a state strictly worse. For a standard approach we get the following level-5 disutilities:  $MO5 = 0.262$ ,  $SC5 = 0.277$ ,  $UA5 = 0.187$ ,  $PD5 = 0.468$ ,  $AD5 = 0.225$ , and the estimated utility  $u(55555) = -0.418$ . Our proposed approach yields 0.222, 0.234, 0.163, 0.423, 0.202, and  $-0.245$ , respectively. Accounting for censoring increases the spread further. The standard approach may overestimate the importance of quality of life (intuitively, a single person with very negative utilities drives the value set down). More discussion is needed on combining utility data from multiple respondents.

### Response Quality in Discrete-Choice Experiments: An Extreme Example of Detecting Fraud

C. Mansfield<sup>1</sup>, J. Sutphin<sup>1</sup>, K. Gallaher<sup>1</sup>

<sup>1</sup>*RTI Health Solutions*

**Background:** Data quality issues in discrete-choice experiments (DCEs) may arise from comprehension problems, inattention to the survey, and outright fraud. We conducted two DCE surveys that were found to contain fraudulent respondents, and we explored whether common methods for assessing data quality can identify fraudulent responses.

**Methods:** Two DCE surveys measuring preferences for treatment of a chronic condition included two standard approaches to identifying potential data quality issues (comprehension questions and a dominated choice). Incorrect responses may indicate a lack of respondent comprehension or inattention but do not explain why respondents answered in unexpected ways. We estimated a random-parameter logit (RPL) model with and without respondents who failed the comprehension and dominated choice questions. A latent class analysis (LCA) model was estimated, which produced multiple classes with intuitive results and classes with disordered results. Subsequently, approximately half the respondents were discovered to be fraudulent data entered by hackers. The data were reanalyzed to identify differences in the responses provided by real and fraudulent respondents.

**Results:** Data quality problems were suspected based on unusual patterns in the demographic variables (fraudulent respondents were more likely to report being male, higher income, and having the chronic condition) and > 50% of respondents failing the comprehension questions. RPL results produced disordered attributes with large confidence intervals. Dropping respondents who failed comprehension and dominated pair questions improved the RPL results marginally. In the two surveys, 23–38% of the fraudulent respondents passed the dominance and comprehension questions, compared to 51–62% of non-fraudulent respondents. In the LCA, fraudulent respondents had a high and significant probability of being in the disordered classes.

**Conclusions:** In this extreme example, patterns in the data suggested unusual data problems. The LCA analysis was reasonably successful in creating classes that distinguished between the preferences of fraudulent and non-fraudulent respondents.

## Comparing Online and Face-to-Face Data Quality and Preferences in a Health Valuation Study

R. Jiang<sup>1</sup>, A. Mühlbacher<sup>2</sup>, J. W. Shaw<sup>3</sup>, T. A. Lee<sup>1</sup>, S. Walton<sup>1</sup>, A. S. Pickard<sup>1</sup>

<sup>1</sup>Department of Pharmacy Systems, Outcomes, and Policy, University of Illinois at Chicago College of Pharmacy, Chicago, IL, USA;

<sup>2</sup>Health Economics and Healthcare Management, Hochschule

Neubrandenburg, Neubrandenburg, Germany;

<sup>3</sup>Patient-Reported Outcomes Assessment, Worldwide Health Economics and Outcomes Research, Bristol-Myers Squibb, Lawrenceville, NJ, USA

**Background:** Online data collection using panels has significant cost and time efficiency advantages over traditional methods of data collection, e.g. face-to-face (F2F). However, the extent to which data quality and elicited preferences may differ between modes is not well characterized. The aim of this study was to compare preference data as elicited using the cTTO and meta-data (e.g., time spent per task, number of trade-offs made) between F2F and online US survey respondents.

**Methods:** The F2F surveys were interviewer-assisted and implemented using the EuroQol Valuation Technology (EQ-VT) with standardized EQ-5D-5L Valuation Protocol 2.0. It was modified for online self-completion with extensive input from experienced researchers. Both modes used the same EuroQol experimental design and employed the same quota sampling for age, gender, ethnicity, and race. All cTTO data were modelled using linear regression with random intercept at the respondent level (RILS). Modes of administration were compared on elicited values; trading behavior, e.g., trading within positive cTTO values only; meta-data; and value set characteristics, e.g., range of scale.

**Results:** Online respondents (n = 501) gave more values clustered at cTTO values of 0 (15.2% vs. 5.3%) and 1 (32.0% vs. 22.2%) and fewer values at -1 (1.0% vs. 13.7%) than F2F respondents (n = 1134). Online and F2F mean elicited cTTO values differed when compared by health state severity (misery score 15: [Online] 0.65 [F2F] 0.25; misery score 25: [Online] 0.41 [F2F] - 0.29). Compared to F2F, more online respondents did not assign the poorest EQ-5D-5L health state (i.e. 55555) the lowest cTTO value ([Online] 41.3% [F2F] 12.2%) (p < 0.001). A higher proportion of online tasks were completed in 3 trade-offs or less ([Online] 15.8% [F2F] 3.7%), (p < 0.001). Mean time spent per task was similar ([Online] 63.3s [F2F] 66.3s). The range of scale for the F2F sample was larger than the online ([Online] 0.600 [F2F] 1.307)

**Conclusions:** Results suggest that data quality was more of an issue when collected online. Online and F2F data provided dramatically different preferences; models estimated with online data provided much smaller disutilities.