ABSTRACTS



12th Meeting of the International Academy of Health Preference Research

ABSTRACTS

12th Meeting International Academy of Health Preference Research



12TH MEETING OF THE INTERNATIONAL ACADEMY OF HEALTH PREFERENCE RESEARCH

Derek S. Brown, Brown School at Washington University, St. Louis, St. Louis, Missouri, USA

Benjamin M. Craig, University of South Florida, Tampa, Florida, USA

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 12th Meeting will be held on Wednesday and Thursday, 23–24 June 2021, and chaired by Ateesha Mohamed and Shelby Reed. Anyone may register via the IAHPR website and attend online via Zoom Webinar[®]. Each of the nine presentations (abstracts below) will have 30 min (12 min for slides, 18 min for discussion) and listed in order of presentation. The abstract submission system closed on 22 April 2021, and these nine abstracts were invited out of 28 submissions based on the ratings of the tenured members. We will not distribute recordings of these presentations. If willing, presenters may distribute their papers in advance and share their slides.

In addition to the podium presentations, the following nine researchers will give elevator talks (4 min for slides; 6 min for discussion; 16:45–18:15 EST on Wednesday 23 June 2021) to introduce themselves and invite collaboration on their ongoing research (listed in order of their talks): Christin Juhnke, Stefan A. Lipman, Andréa Libório Monteiro, Maksat Jumamyradov, Nicholas Smeele, Eline Bouwers-Beens, Riccarda Peters, Norah L. Crossnohere and Jody L. Church. Like the abstracts, their applications were selected based on the ratings of the tenured members, and we welcome these brief introductions (similar to a virtual poster session).

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.

Session 1 (15:00–16:30 EST, Wednesday 23 June 2021):

The Faster, the Better: QALY Distortions in Valuing Speed of Onset of Action

F. Reed Johnson¹, Juan Marcos Gonzalez¹, Shelby Reed¹

¹Duke University, Durham, NC, USA

Abstract: The QALY assumption that health preferences are separatively additive in a sequence of health states is inconsistent with diminishing marginal utility, which predicts that the quality-adjusted value of a sequence of health improvements depends on timing. We show under what conditions the sum of linear health-state utilities (HSUs) can distort nonlinear time-equivalent values (TEVs) and provide examples of short-term (≤ 1 year) nonlinearity from published DCE studies.

Linear distortion depends on the combined effect of degree of curvature in the utility function and the pattern of clinical health-state improvements over the treatment period. We vary the curvature of a utility function of the CARA form:

$$U(t) = \frac{1 - e^{-at}}{a}a > 0$$

U(t) is linear for a=0. We compare 3 utility-function curvature patterns with 3 possible treatment-effect timing patterns having the same average slope (and thus same QALYs): linear, initial slow, and initial fast speeds of onset of action.

The TEVs for combinations of timing-specific marginal treatment effects and marginal-utility weights are calculated as:

$$TEV_{mn} = \sum_{t=0}^{T} \frac{\Delta U_{nt}}{\Delta H_{mt}} \frac{\Delta H_{mt}}{\Delta t}$$

where H_{mt} is the health index for treatment-effect timing pattern m in period t, U_{nt} is the utility of a utility function with curvature n in period t, and T is the number of treatment periods. We identify cases where differences in timing of health-status improvements can result in linear-additive QALYs overstating, understating, and equaling nonlinear TEV values, depending on how health changes and nonlinear-curvature weights interact.

Unlike QALYs, general TEVs do not require eliciting time/quality trade-off preferences in one-year longevity increments. DCEs increasingly are replacing traditional methods for estimating HSUs. They also offer a tractable means of obtaining evidence on timing-specific, short-term, nonlinear health utility. We discuss results from three published empirical DCE applications that identify diminishing marginal utility in timing for short-duration spells of ill health: migraine (24 h), EQ-5D (2 months), and Crohn's disease (12 months).

References:

[1] Fechner GT. Elements of psychophysics, 1860. In: Readings in the history of psychology. East Norwalk, CT, US: Appleton-Century-Crofts; 1948:206–213.

[2] Bala M, Zarkin G. Are QALYs an appropriate measure for valuing morbidity in acute diseases? Health Econ. 2000;9(2):177–180.

[3] Johnson FR, Hauber B, Ozdemir S. Conjoint-analysis QALYs for acute conditions. Value in Health. 2006;9(6):A258–A258.

[4] Gonzalez JM, Johnson FR, Runken MC, Poulos CM. Evaluating migraineurs' preferences for migraine treatment outcomes using a choice experiment. Headache. 2013;53(10):1635–1650.

[5] Craig BM, Rand K, Bailey H, Stalmeier PFM. Quality-Adjusted Life-Years without Constant Proportionality. Value Health. 2018;21(9):1124–1131.

Comparing Discrete Choice Experiment (DCE) & Probabilistic Threshold Techniques (PTT) in 3 Countries

Rachael L. DiSantostefano¹, Marie Falahee², Gwenda Simons², Matthias Englbrecht³, Christine Radawski⁴, Ellen Janssen¹, Brett Hauber^{4,5}, Karim Raza², Jorien Veldwijk^{6,7}

¹Janssen Research & Development, Titusville, NJ, USA; ²Rheumatology Research Group, Institute of Inflammation and Ageing, College of Medical and Dental Sciences, University of Birmingham, Birmingham, UK; ³Eckental, Germany; ⁴Eli Lilly & Company, Indianapolis, Indiana, USA; ⁵Pfizer, New York, NY, USA; ⁵Comparative Health Outcomes, Policy, and Economics (CHOICE) Institute, University of Washington School of Pharmacy, Seattle, Washington, USA; ⁶Erasmus School of Health Policy & Management and Erasmus Choice Modelling Centre, Erasmus University Rotterdam, The Netherlands; ⁷Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht University, Utrecht, The Netherlands

Background: DCE studies are commonly applied to elicit preferences but are perceived as being difficult for respondents. This raises the question about whether a less complex approach can give similar results. The main objective of this study was to compare two statedpreference methods, DCE and PTT, to assess differences in preferences measured and ease of use of the two methods [1].

Methods: In the United Kingdom, Germany, and Romania (n = 2959), representative samples of the general public completed a DCE and PTT in random order to elicit preferences for treatments that reduce a baseline chance of getting rheumatoid arthritis (RA) within 2 years. For each country separately, random parameters logit (RPL) models (DCE) and interval regression models (PTT) were used. Model results (relative importance, maximum acceptable risk [MAR]) and participant feedback were compared across methods.

Results: Results were consistent across countries. For a 40 percentage point reduction in chance of developing RA (60–20%), MARs did not differ between methods for serious infection or serious side effects but were dissimilar for mild side effects (e.g., 45.8% DCE v. 15.8% PTT in UK). The majority found both methods easy/very easy to complete with the DCE reported being easier (p < 0.05). Respondents who completed the DCE first found both methods easier to understand and easier to answer (p < 0.05).

Conclusions: Across all countries, MARs for the two relatively more important attributes did not differ across methods, but MAR differed for the relatively less important attribute. Both DCE & PTT were easy for a majority of participants to understand and complete, with DCE being easier. DCE has an advantage of a multi-attribute approach that considers all trade-offs simultaneously. However, PTT might be equally suitable when considering more important attributes, with a simpler design and fewer questions.

References:

[1] Falahee M, et al. Treatment preferences for preventive interventions for rheumatoid arthritis: protocol of a mixed methods case study for the Innovative Medicines Initiative PREFER project. BMJ Open 2021;11:e045851. https://doi.org/10.1136/bmjopen-2020-045851

Can We Move Beyond First-Choice DCEs Using Best-Worst Using Best-Worst, Best-Best or Ranking?

Samare P.I. Huls¹, Emily Lancsar², Bas Donkers¹, Jemimah Ride³

¹Erasmus University Rotterdam, The Netherlands; ²Australian National University, Canberra, Australia; ³University of Melbourne, Melbourne, Australia

Background: In addition to asking respondents their first choice, there is an increasing interest to ask them to choose among the remaining alternatives of a choice set in Discrete Choice Experiments (DCEs). To inform the decision on which preference method to use if wanting to move beyond traditional first-choice DCE, this study makes a head-to-head comparison of best-worst, best-best and ranking discrete choice experiments.

Methods: The study consisted of three arms, respondents were randomised to one. Each arm involved an identical experiment and only differed in the elicitation method: best-worst; best-best and ranking. The three methods were compared using six criteria: trade-off consistency, choice consistency, scale dynamics, efficiency, stated difficulty and stated preference. Rank-ordered mixed logit models and respondent-reported data were used to compare the criteria between arms and first and second choices.

Results: Choices were most consistent in ranking and least consistent in best-worst, especially the second "worst" choice. Learning effects and efficiency were largest for best-worst, especially in the second choice. Furthermore, ranking was perceived to be easiest and most preferable. However, trade-offs differed more by first or second choice within a preference elicitation method than between the three methods. Respondents were more consistent in first choices, these were also reported to be easier.

Conclusions: All methods improve efficiency of data collection relative to using first choices only. However, even after allowing for differences in scale and scale dynamics, first choices reflect preferences that differed from those of second choices for all three preference elicitation methods. This raises doubts whether to move beyond first-best choice preference elicitation methods.

Session 2 (7:00-8:30 EST, Thursday 24 June 2021):

Should We Continue to Include Repeated and Dominant Tasks in Discrete Choice Experiment Designs?

Marcel F. Jonker¹, Bram Roudijk², Merit Maas¹

¹Erasmus University, Rotterdam The Netherlands; ²EuroQol Research Institute, Rotterdam, The Netherlands

Background: Although there is a growing consensus on good practice in several areas of health preference research, there are no widely accepted indicators of response quality in discrete choice experiments (DCEs).

Methods: A targeted review of recent DCE articles (2018–2020Q1) published in the health, marketing, and transport economics literature was used to identify the most commonly used internal validity tests. Two of these were then incorporated in four different (online) data collections. Based on the estimated respondent preferences, the achieved sensitivity and specificity was simulated and compared with the sensitivity and specificity of the most commonly used (RLH) statistical test.

Results: Dominant and repeated choice tasks are included in about 1/5 of the recent DCE publications and by far the most commonly

used internal validity tests. Across different datasets, their sensitivity ranges from 76 to 83%, while their specificity depends on the type of invalid response pattern (e.g. random, deterministic, etc) but ranges from 32 to 68%. In comparison, the RLH test was found to have a sensitivity of 79–97% and specificity of 46–94%, of which 93–94% for random response patterns.

Conclusions: Dominant and repeated choice tasks are relatively unreliable at identifying high and low-quality respondents as well as costly to include in terms of statistical power. The root-likelihood test, in contrast, does not require additional choice tasks to be included in the DCE design and was found to provide a superior alternative that never performs worse and often performs substantially better, particularly at identifying random response patterns.

A Comparison of Full and Partial Choice Set Designs in a Labelled Discrete Choice Experiment

Thao Thi Hong Thai¹, Michiel Bliemer², Gang Chen¹, Jean Spinks³, Sonja de New¹, Emily Lancsar⁴

¹Monash University, Victoria, Australia; ²The University of Sydney, New South Wales, Australia; ³Griffith University, Queensland, Australia; ⁴Australian National University, Canberra, Australia

Background: One approach to reducing complexity of large labelled choice tasks is to use a partial choice set design (PCSD) in which a subset of alternatives is shown in each choice task in contrast to a traditional full choice set design (FCSD) where all alternatives are shown. Using a within-respondent comparison, we aimed to (1) test the convergent validity of the PCSD and FCSD; (2) verify if the PCSD reduces cognitive burden; (3) explore respondents' preferences between the FCSD and PCSD.

Methods: Labelled utility functions were rewritten into a single generic utility function using a label dummy variable and indicator functions, which was used to create a PCSD with 3 alternatives in each task (out of 6). The convergent validity of two designs' results from conditional logit and mixed logit models were tested using the Swait and Louviere test and the convolution test, respectively. The PSCD's impact on choice variances was examined using a heteroscedastic conditional logit model.

Results: Using data from 790 respondents, we found preference estimates from the FSCD and PSCD are statistically different up to scale. These results remain different even accounting for more flexible substitution patterns across alternatives using the MIXL model. We found that the PCSD appeared to induce smaller choice variance than the FCSD, which reflects positively on its purpose of reducing the cognitive burden. The PSCD was preferred by female and when phones were used to answer the survey.

Conclusions: Our findings indicate that the PCSD can reduce cognitive burden and we suggest its use for surveys accessible by mobile phone. While both a PCSD and FCSD should capture the same behaviour, our study reveals statistically significant differences, perhaps because respondents in a FCSD were not trading off on all attributes and alternatives due to choice task complexity, but without more research on external validity it is not possible to conclude which design type better uncovers true preferences.

References:

[1] DeShazo, J.R., & Fermo, G. (2002). Designing Choice Sets for Stated Preference Methods: The Effects of Complexity on Choice Consistency. Journal of Environmental Economics and Management, 44, 123–143.

[2] Lancsar, E., & Louviere, J. (2006). Deleting 'irrational' responses from discrete choice experiments: a case of investigating or imposing preferences? Health Econ, 15, 797–811.

[3] Rose, J.M., Louviere, J.J., & Bliermer, M.C.J. (2013). Efficient stated choice designs allowing for variable choice set sizes. International Choice Modelling Conference.

[4] Bliemer, M., Rose, J., & Matthew, B. (2018). Generating partial choice set designs for stated choice experiments 15th International Conference on Travel Behaviour Research Santa Barbara.

[5] Swait, J., & Adamowicz, W. (2001a). Choice Environment, Market Complexity, and Consumer Behavior: A Theoretical and Empirical Approach for Incorporating Decision Complexity into Models of Consumer Choice. Organizational Behavior and Human Decision Processes, 86, 141–167.

Improved External Validity of DCE Uptake Predictions Based on a Dual-Response None Option Format?

Marcel F. Jonker¹, Esther W. de Bekker-Grob¹, Jorien Veldwijk¹, Lucas Marcus Anthonius Goossens¹, Maureen Rutten-Van Molken¹

¹Erasmus University Rotterdam, The Netherlands

Background: A dual-response none option format can provide (potentially useful) preference information for respondents who otherwise consistently choose the opt-out option. Unfortunately, there is still a lack of evidence about the quality of these additional choice data. Morover, the different framing of the opt-out options could very well induce slightly different choice behavior, resulting in different preference estimates and potentially very different uptake predictions. **Methods:** To investigate the impact of the opt-out elicitation format on preference estimates, uptake predictions, and data quality, an existing COVID-19 instrument was re-fielded using a standard and dual-response none option format. This resulted in two nationally representative samples of approx. 1,000 respondents each. These data were analyzed using raw data tabulations, MIXL (individual-level) uptake predictions, and data quality was determined using statistical root-likelihood (RLH) tests.

Results: In the standard none sample, 24% always, 26% sometimes, and 50% never chose the opt-out, resulting in a 60% predicted uptake. In the dual-response sample, 28% always, 37% sometimes, and 36% never chose the optout, resulting in a 49% predicted uptake. Compared to the external benchmark of 35-40%, the dual-response format thus resulted in better external validity. Data quality was also better in the dual-response sample with fewer respondents identified as having used a random response pattern.

Conclusions: The dual-response none format resulted in improved external validity without an obvious trade-off in terms of data quality - a conclusion that also holds for respondents who consistently choose the none option. These respondents did appear to have distinctively different preferences, which implies that practioners have to be cautious when pooling their choice data in a single (combined) statistical model.

Session 3 (8:45–10:15 EST, Thursday 24 June 2021):

Modelling the Impact of Benefit and Risk Attributes in Health DCEs: The Role of Attribute Screening

Jorien Veldwijk¹, Joffre Swait¹

¹Erasmus School of Health Policy and Management, Erasmus University Rotterdam and Erasmus Choice Modeling Centre, Erasmus University Rotterdam, The Netherlands **Background:** Besides the lack of evidence-based guidance on accurate health-risk communication in DCEs, modelling the impact of benefit/risk attributes on the preferences of respondents merits further deliberation. Screening or elimination of alternatives before choice (i.e, choice set formation [1, 2]) might be used in health related DCEs that include benefit-risk attributes. This study aims to demonstrate the econometric modelling of benefit/risk-based choice set formation within health-related DCEs.

Methods: In four different case studies first a standard trade-off model was fitted (multinomial logit model), building on this a screening model was fitted and finally a full choice set formation model was estimated. This final model allows for attributes to be used first to screen out alternatives from choice tasks before respondents' trade-off attributes and make a choice among feasible alternatives [1, 2]. Educational level and health literacy of respondents was accounted for in all models.

Results: Model fit in (e.g, Log Likelihood & BIC) improved from using only trade-off or screening models compared to choice set formation models in three out of four studies. In those studies, significant screening behavior was identified which significantly impacted trade-off inferences, rejecting the pure trade-off model and supporting the existence of screening on the basis of benefit/risk profiles. Educational level and health literacy showed significant interactions with attributes in all studies.

Conclusions: Choice modelers should pay close attention on how respondent behave when they include benefit/risk attributes in their DCE. Further studies should investigate why and when respondents undertake screening behavior. Researchers should explore extensions of econometric models to reflect non-compensatory behavior. Assuming benefit and risk attributes will only impact trade-off behavior is likely to lead to false conclusions about benefit/risk-based behavior.

References:

[1] Swait, J., Benakiva, M., Incorporating Random Constraints in Discrete Models of Choice Set Generation. Transportation Research Part B-Methodological, 1987. 21(2): p. 91–102.

[2] Swait, J., Choice set generation within the generalized extreme value family of discrete choice models. Transportation Research Part B-Methodological, 2001. 35(7): p. 643–666.

Health Baselines and Stated Preferences: Quantifying the Evolution of Preferences Using LCA

Juan Marcos González Sepúlveda¹, Jui-Chen Yang¹, Shelby D. Reed¹, Ting-Hsuan Lee², Sarah Stothers², Kaiwen Li¹, Xinyi Ng², Martin Ho², Telba Irony³

¹Duke Clinical Research Institute, Durham, North Carolina, USA; ²United States Food and Drug Administration, Center for Biologics Evaluation and Research, Silver Springs, Maryland, USA; ³Janssen Pharmaceutical, Raritan, New Jersey, USA

Background: Accurate evaluation of treatment preferences in DCEs require defining a clear decision context that asks respondents to consider self-reported or a standardized health baseline. Standardized baselines fix the context and help ensure trade-off plausibility, but also presents a double hypothetical to respondents as they make their choices. Our objective was to compare, at the respondent level, how treatment preferences varied when respondents used their own baseline versus more severe baselines.

Methods: A DCE survey was administered to patients and parents of minors (< 18) with sickle-cell disease to evaluate willingness to pursue gene therapy. Respondents answered choice questions for both their self-reported baseline and an assumed, more severe, baseline.

Latent classes were identified by cohort and baseline status. Changes in individual respondent's class-membership probabilities by actual and hypothetical baselines were used to evaluate how baseline framing affected preference estimates.

Results: 174 patients and 109 parents completed the survey. Multiple classes were identified across cohorts by baseline. Class allocation was above 93% for all classes and progressed unevenly with standardized baselines. The relative value of opting out of gene therapy changed with baselines, yet this value varied less across self-reported baselines. A significant difference in preferences found between patients with moderate symptoms and those assuming moderate symptoms was not observed among parents.

Conclusions: The analysis of class progression can be an effective tool to evaluate how individual respondents react to changes in choice contexts. Our results indicated that respondent choices accounted for varying baselines. Adaptive behaviors seem plausible given unchanged values for opt out among some respondents with current mild symptoms and some with current moderate symptoms. Overall, we found some support for the use of standardized baselines to represent clinically-relevant choice contexts.

Matching and Weighting in Health Preference Research

Caroline M. Vass^{1,2}, Marco Boeri^{3,4}, Christine Poulos⁵, Alex J. Turner^{1,6}

¹RTI Health Solutions, Manchester, UK; ²Manchester Centre for Health Economics, University of Manchester, Manchester, UK; ³RTI Health Solutions, Belfast, UK; ⁴Queens University Belfast, Belfast, UK; ⁵RTI Health Solutions, North Carolina, USA; ⁶Health Organisation, Policy and Economics Group, University of Manchester, Manchester, UK

Background: There is a growing interest in quantifying the degree of heterogeneity in stated preferences for health. A popular investigation into preference heterogeneity involves split-sample analysis to make comparisons across subgroups. However, subgroups may differ in many observed characteristics. Not accounting for these other characteristics may bias comparisons if these are also associated with preferences. This study explores matching and weighting approaches to identify differences in preferences.

Methods: We compare simulated stated preferences of patients and the public for a hypothetical healthcare intervention, where patients are older and have lower household income. The utility function for both is specified to be identical (preference homogeneity) and utility is assumed to increase with health and life years, and decrease with risk and cost. Utility for cost is specified as a function of income and age. We conduct unmatched, propensity score-matched, and entropy balanced analyses.

Results: Due to differences in age and income, unmatched analysis detects statistically significant differences in the preference for cost when comparing the public's preferences with those of patients. Both propensity score matching and entropy balancing reduce imbalance in the individual characteristics across subgroups, although the reduction is greater when using entropy balancing. Following matching or weighting, there are no significant differences in the preference weights for any attributes.

Conclusions: Unweighted and unmatched analyses may produce erroneous conclusions regarding heterogeneity in preferences when making comparisons across subgroups. Matching and weighting methods may be useful for researchers seeking to compare preferences for health and health care when there are too many characteristics to feasibly incorporate with interaction terms.