THE INTERNATIONAL ACADEMY OF HEALTH PREFERENCE RESEARCH
IS PROUD TO ANNOUNCE

THE 3RD MEETING OF THE INTERNATIONAL ACADEMY OF HEALTH PREFERENCE RESEARCH

SUNDAY, OCTOBER 18, 2015
8:00 AM - 5:30 PM

Chaired by Derek Brown, PhD, Washington University in St. Louis

CHARLES F. KNIGHT EXECUTIVE EDUCATION & CONFERENCE CENTER
AT WASHINGTON UNIVERSITY IN ST. LOUIS
ONE BROOKINGS DRIVE
ST. LOUIS, MISSOURI, USA

Held at the Knight Center—located on the grounds of Washington University—this 1-day meeting will provide a forum to discuss innovative developments in the field of health preference research. Chaired by **Derek Brown, PhD**, the meeting will include approximately 13 oral presentations, light breakfast, coffee, lunch, and a business session. All are welcome to **register at:** www. iahpr.org.

PRE-MEETING DINNER & STUDENT POSTER SESSION
SATURDAY, OCTOBER 17, 2015 – 6:00 TO 11:30 PM

The Pre-Meeting Dinner and Student Poster Session will also be held at the Knight Center and is free for all meeting attendees. The dinner includes a multi-course menu and open bar (no guests, please). The poster session was created as a way to showcase the achievements of students engaged in health preference research. The event includes a free shuttle between the Hyatt Regency and Knight Center. The shuttle schedule will be included in the registration packet.

For more information, visit www.iahpr.org **or email** meeting2015@iahpr.org .

PROGRAM

Pre-Meeting Dinner & Student Poster Session, Saturday, October 17, 2015 – 6:00 to 11:30 PM

Investigating the Impact of Individual Valuation Block Composition on TTO Estimates
Andréa Libório Monteiro [β]
Specialist Training as an Incentive to Retain Doctors in Malawi: A Discrete Choice Experiment
Kate Mandeville [β]
Comparison of PROMIS and EQ-5D Quality-Adjusted Life Years
John D. Hartman [β]
Stated-Preference Survey Development for Muscular Dystrophy: A Community-Engaged Research Application
Ilene L. Hollin [β]

Meeting, Sunday, October 18, 2015 – 8:00 AM - 5:30 PM

8:00-8:30 AM	**Arrival and Light Breakfast**
8:30-8:45	**Welcome and Acknowledgement of Sponsors** **Meeting Chair: Derek Brown**

8:45-10:15 **Session 1**
What if You Ask Them Again? Temporal Stability of Choices and Preference Estimates from a DCE
Jan Ostermann [α]
Radial versus Femoral Vascular Access Options in Coronary Angiography and Intervention
Janine van Til
Willingness-to-Pay for Health: A Fuzzy Approach to Modelling Preferences and Choice Functions
Michał Jakubczyk
Best-Worst Scaling Works with Virtually Everyone. Except Kids
Terry Nicholas Flynn [α]

10:15-10:30 **Coffee Break**

10:30-12:00 PM **Session 2**
Use of Best-Worst Scaling to Assess Patient Perceptions of Refractory Overactive Bladder Treatments
Kathleen Marie Beusterien [α]
Valuing New HRQOL Measures: A DCE Application for Adverse Childhood Experiences and Maltreatment
Derek Brown [α]
The Effect of Framing of Death on Health State Values Obtained from Discrete Choice Experiments
Marcel Jonker
Framing of Attribute's Levels: Influence on the Interpretation of Outcomes from a BWS Experiment
Marieke Weernink [β]

12:00-1:00 **Lunch**

1:00-2:30 **Session 3**
Attribute Non-Attendance and Time Pressure in Discrete Choice Experiments: An Eye-Tracking Study
Kate Mandeville [β]
Respondent Cognition in Health Preference Research
Shannon K. Runge [β]
Mitigating Hypothetical Bias in Stated Preference Discrete Choice Experiments
Dean A. Regier [α]
Discount Rate Assessment among Adults Experiencing Dyspnea from Common Primary Care Diseases
Irene D. Fischer

2:30-2:45 **Coffee Break**

2:45-3:45 **Session 4**
Conjoint Analysis: A Tool for Understanding Patients Decisions for Invasive Treatments
Tracy Kuo Lin
Open Discussion
Derek Brown

4:00-5:30 **Business Meeting (All attendees are welcome)**

[α] indicates a member presenter
[β] indicates a student presenter

The 3rd Meeting of the International Academy of Health Preference Research

---

Table of Contents

Benefactors

Sponsors

8:45-10:15 AM   Session 1

What if You Ask Them Again? Temporal Stability of Choices and Preference Estimates from a DCE

Jan Ostermann, PhD, Duke University; *Derek Brown, PhD,* Washington University in St. Louis; *Axel Muehlbacher, PhD,* Hochschule Neubrandenburg; *Nathan Thielman, MD, MPH,* Duke University

Purpose: Discrete choice experiments (DCEs) are increasingly used to estimate preferences for diverse health-related goods and services. Preference estimates are usually derived from a single assessment, but tend to be used to inform sometimes much delayed, longer-term resource allocation decisions and policies. The temporal stability of DCE-based preference estimates is critical for the method's utility for informing policy design and implementation. This study sought to assess the stability of individual respondents' choices and aggregate DCE-based preference estimates over time.

Methods: Between September 2012 and February 2013, a DCE on HIV testing preferences was conducted among 486 randomly selected community members in an urban setting in Northern Tanzania. The DCE contained 5 attributes: distance to testing, confidentiality, testing days (weekday vs. weekend), method for obtaining the sample for testing (blood from finger or arm, oral swab), and availability of HIV medications at the testing site. An NGene generated D-optimal design included 72 forced choice tasks, allocated over 8 blocks. Each participant was given 9 choice tasks assigned in random order. The DCE was implemented on iPads. To assess the temporal stability of individual choice patterns and aggregate preference estimates, participants were re-contacted after an average of 8.2 months (sd=1.5 months) and asked to repeat the exact same choice tasks in the exact same order. Gender-specific mixed logit models were used to estimate mean preference parameters; interactions with an indicator variable for the repeat assessment were used to evaluate the stability of preference estimates over time. Ordered probit models evaluated the effect of HIV testing experience, HIV serostatus disclosure, and changes in risk characteristics between the two assessments on the number of changed responses.

Results: In total, 301 participants repeated the 9 choice tasks. On average, participants changed their answers on 3.2 of 9 choices. Failure to provide the same answer on repeat assessment was associated with a first-time HIV test between the two assessments, HIV serostatus disclosure to spouse and others, and incident HIV risk characteristics, including commercial sex and more than one sexual partner between the assessments. Gender-stratified estimates of average preferences were not significantly different between assessments for 14 of 18 attribute levels, and rankings of relative preference weights across all 14 attribute levels differed by <2 ranks, on average.

Conclusions: In this context and population, the estimated mean preference parameters derived using DCE methods have good temporal stability. Changes in respondent characteristics are associated with changed responses in DCE choice tasks, providing additional support for the validity of DCE methods for eliciting preferences.

Radial versus Femoral Vascular Access Options in Coronary Angiography and Intervention

Anneloes Fens, MSc, and *Marieke Weernink, PhD student,* University of Twente; *Clemens von Birgelen, MD, PhD,* Medisch Spectrum Twente, Enschede; *Janine van Til, PhD,* University of Twente

Purpose: For years, the trans-femoral access (TFA) was the default access route to the coronary arteries because the TFA is a fail save and easy access route from a clinician's perspective. In recent years, there is an increasing interest in trans-radial access (TRA) as it has advantages that mainly benefit the patient, such as earlier patient ambulation and in increased comfort. Success rates of TFA and TRA itself are comparable. The objective of this study is to determine patient preferences over benefits, risks and process characteristics of coronary intervention and to relate them to patient previous experiences.

Methods: A Best-Worst Scaling (BWS) case 2 was conducted to elicit patient preferences (n=153) for six attributes of care: risk of bleeding, need to change access-site during the procedure, vessel quality post

procedure, length of hospital stay, post-procedural patient comfort, and post-procedural mobilization. Patients were asked to indicate which attribute of treatment they perceived as the most and lease important in choosing for treatment, the results being equal. Moreover, patient's previous experience with treatment and direct stated preference for treatment were elicited. Patients were provided with an (oral) explanation by a researcher. Best-minus-Worst scores, conditional logit analysis, attribute importance, and subgroup analysis were calculated and/or executed.

Results: Overall, patients (n=143) considered the risk of bleeding as most important attribute of treatment (attribute importance 31%), followed by length of hospital stay (23%) and post-procedural mobilization (20%). When direct preferences for access route were elicited, a small majority of patients (60%) preferred the femoral route. Many patients had an intervention prior to the current intervention. Within the subgroup of patients that had a history including both vascular routes, 72% were in favor of the radial route. Subgroup analysis of attribute importance revealed that patients who had prior interventions, attached more importance to hospital stay (28 vs. 18%) and less importance to risk of bleeding (10 vs. 30%) compared to patients without a prior intervention. No major differences in preference in attribute importance were found between patients that currently had femoral or radial access.

Conclusion: While lower risk of bleeding, shorter length of hospital stay and quicker post-procedural mobilization were all in favor of radial access route, the majority of patients were still in favor of femoral access when asked directly. Previous experience has a major impact on direct treatment preference. Given that because of organizational reasons, patients were interviewed after treatment took place, this might have been expected. We propose using preference elicitation as a tool to enhance shared decision making in the cardiology department. However, this includes balancing patient and clinician's preference.

Willingness-to-Pay for Health: A Fuzzy Approach to Modelling Preferences and Choice Functions

Michał Jakubczyk, PhD, Warsaw School of Economics

Purpose: Allocating public resources between health technologies, to be socially optimal, requires health to be first defined and measured (often using preference-based approach, e.g., using quality-adjusted life years), and then valued in monetary terms. The latter is difficult, e.g., due to health being a non-market good (opposite to health services) and strong ethical issues involved; hence, the willingness-to-pay (WTP) often cannot be precisely determined. The of this research is to advocate using fuzzy sets to modelling WTP and to show how this approach can be used to model decision makers preferences and to support decision making via fuzzy choice functions.

Methods: A preliminary survey among 27 health technology assessment (HTA) experts in Poland was performed to i) verify the presence of fuzziness in valuing health, ii) detect possible discrepancies (suggested in the literature) between the WTP and the willingness-to-accept (WTA), iii) compare different elicitation methods (Likert-based vs direct asking for a range of values). An axiomatic approach to define the structure of preferences involving two criteria: health and money: was suggested. Fuzzy counterparts of notions typically used in HTA were suggested: fuzzy net benefit and fuzzy (cost-effectiveness) acceptability curve in case uncertainty is present. A fuzzy choice function was defined when the choice is made from among possibly more than two alternatives. Properties of newly defined concepts were verified.

Results: A majority of experts agreed that a threshold value of health should be determined and publicly known in HTA decisions. At the same time a majority could only report their preferences in a fuzzy manner. Likert-type questions reveal more nuanced information on the WTP/WTA: asking directly about the WTP range makes respondents focus only on values of which they are quite convinced. The WTP/WTA seem to meet the standard properties of fuzzy preferences (connectedness, antisymmetry), and so the WTP/WTA disparity results rather from the fuzzy perception of both and above-mentioned focusing on ranges of which the respondent is highly convicted. Fuzzy net benefit respects dominance and extended dominance relation between health technologies. Fuzzy acceptability curves in some cases have more intuitive properties than a non-fuzzy original (which need not to be monotonic). Fuzzy choice function can be defined in case several alternatives are available, but the actual choice requires crispification as the last stage. Defuzzification via maximizing the conviction that an option is the best (i.e., selecting the option

with greatest membership) may violate the alpha-property of the choice function. Accounting for the natural ordering of decision alternatives (e.g., by cost) may lead to a different, median-based defuzzification, which removes the alpha-violation and also WTP/WTA disparity.

Conclusions: Using fuzzy set theory is a natural approach to thinking about preferences regarding the cost vs effectiveness trade-offs. This fuzziness can be defined axiomatically, elicited, and introduced formally into decision making. Fuzzy approach allows to explain some phenomena, e.g., WTP/WTA disparity, and this explanation can also be used in a crisp setting. Fuzzy thinking can also be introduced into other areas of modeling health preferences.

Best-Worst Scaling Works with Virtually Everyone. Except Kids

Terry Nicholas Flynn, PhD, MSc, BA, Tf Choices LTD (UK); Elisabeth Huynh, PhD, B. Com, University of South Australia; Gang Chen, PhD, and Julie Ratcliffe, PhD, Flinders University

Purpose: The dominant paradigm in health economic evaluation requires a representative sample of the wider general adult population to value impaired health states for a standardized health instrument (such as the EQ-5D). The resulting set of values: a "tariff": is then used in economic evaluation to value effectiveness. Unfortunately, previous pilot work with the CHU9D child health instrument suggested that adults under-weight mental health related dimensions compared to adolescents. The purpose was to establish if this is true in a large sample study.

Methods: Aims - 1. To provide an Australian adolescent population tariff for the CHU9D 2. To demonstrate how any heterogeneity in preferences manifests itself among adolescents 2020 Australian adolescents aged 11-17 were recruited via their parents, who were members of the pure profile online panel. A Case 2 Best-Worst Scaling (BWS) exercise was administered to estimate the relative disutility of all attribute levels described by the CHU9D. Respondents faced ten CHU9D health states (one fifth of the master experimental design in 50 : the near-orthogonal main effects plan used in the pilot) and asked to consecutively indicate the best, worst, second best and second worst attributes (dimension levels) of each presented state. This was intended to provide more power to heterogeneity analyses if these four ranks could be pooled. The BWS task was followed by a number of socio-demographic and health questions. Analysis used conditional logit models and their extensions to estimate the disutility associated with various CHU9D impairments and latent class analyses to check for preference heterogeneity and test whether data from the four rankings (best, second best, second worst, and worst) could be pooled.

Results: Examination of various latent class solutions, conducted on all four sets of ranking data, showed differences by rank: essentially, unlike the assumption made in classical ranking estimators (such as the rank ordered logit), different utility functions are obtained at different ranking depths. Furthermore choice consistency was noticeably poorer at the three non-first-best ranks. This, together with examination of the utility estimates, led to the rejection of the lower three sets of ranking data and the estimation of a tariff using the first best data only. Two latent classes were found, one (around two thirds of respondents) placing strong weight on mental health impairments whilst the other seemed largely to ignore lower levels of all attributes.

Conclusions: Adolescents do indeed believe mental health related impairments described by the CHU9D to be worse than adults. This means that interventions here may appear worse in economic evaluations if the current practice of using adult values is maintained. Since Sweden prefers values to come from the individuals affected (patients), Swedish valuation of child health should be a research priority. BWS is rapidly being adopted worldwide to elicit preferences, given how well it performs among vulnerable groups like the elderly, carers and patients. However, this study is the first to show that adolescents are one group for whom it does not work well, and conventional discrete choice experiments (requiring "first choice" only) should continue to be used.

Use of Best-Worst Scaling to Assess Patient Perceptions of Refractory Overactive Bladder Treatments

Kathleen Marie Beusterien, MPH, *Outcomes Research Strategies Health, Washington DC;* **Michael Kennelly, MD,** *Women's Center for Pelvic Health, Charlotte, NC;* **John F. P. Bridges, PhD,** *Johns Hopkins University, MD;* **Kaitlan Anne Amos, BS,** *Outcomes Research Strategies Health;* **Mary Jo Williams, BS,** *Medtronic;* **Sandip Vasavada, MD,** *Cleveland Clinic, OH*

Purpose: Current treatments for refractory overactive bladder (OAB) differ considerably: sacral neuromodulation (SNM) involves surgically implanting a device in the upper buttock; onabotulinumtoxinA involves injections into the bladder muscle; and percutaneous tibial nerve stimulation (PTNS) involves inserting a needle into the ankle to stimulate the tibial nerve. As such, a traditional discrete choice experiment is impractical for assessing attribute preferences in this scenario. This study used Best-Worst Scaling (BWS) case1 to assess patient perceptions of these treatments.

Methods: A cross-sectional Web survey, based on findings from qualitative interviews with 23 OAB patients and 7 clinicians, was conducted with idiopathic OAB patients in the US and UK. The BWS exercise involved prioritizing subsets of 13 attributes across 13 choice tasks, where patients identified the attribute they considered as best and worst in each task. Attribute BWS scores, ranging from -1.0 (worst) to 1.0 (best) were calculated based on the rates that each attribute was chosen. Attitudes toward the attributes also were assessed via like/dislike Likert scales, and patients were asked their percentage likelihood (0-100%) of trying each treatment, based on standardized treatment descriptions.

Results: 245 patients (118 US, 127 UK) completed the survey (79% female; mean age of 50 + 7.8). All 13 attribute BWS scores were significant, reflecting that patients viewed them positively (score > 0.0) or negatively (score < 0.0). "Lasting improvement" (0.82), "Minimal side effects" (0.67), and "Sends signals between bladder and brain" (0.35) were rated most favorably, and "Complications with implant" (-0.65), and "Be willing to self-catheterize" (-0.53) were rated worst. The percentage likelihood estimates for trying one of the three treatments were significantly correlated with all of the BWS scores except one ("MRI restriction"). Specifically, the likelihood of trying SNM was correlated with "implanted device" (r=0.49) and "sends signals" (r=0.34), and negatively correlated with "repeated visits" (r=-0.35), "needle in ankle" (r=-0.27) and "minimal side effects" (r=-0.25). The likelihood of trying onabotulinumtoxinA was correlated with "Botox (botulinum toxin) treatment" (r=0.53), "self-catheterize" (r=0.30), "treatment via urethra" (r=0.25), and "minor procedure" (r=0.16), and was negatively correlated with "needle in ankle" (r=-0.29), "implant complications" (r=-0.27), "repeated visits" (r=-0.21), "implanted device" (r=-0.18), and "test phase" (r=-0.15). The likelihood of trying PTNS was correlated with "needle in ankle" (r=0.27), "sends signals" (0.20), and "lasting improvement" (r=0.14), and was negatively correlated with "minor procedure" (r=-0.21) and "Botox (botulinum toxin) treatment" (r=-0.13). Although all but one like/dislike score correlated significantly with the respective BWS scores, the correlations between like/dislike scores and willingness to try each treatment differed substantially from the BWS score-willingness to try treatment correlations. Specifically, all like/dislike-willingness to try treatment correlations were positive, thus disliking an attribute did not translate into significant correlations with willingness to try an alternative treatment, as was observed with the BWS scores.

Conclusions: BWS was successful in assessing the magnitude of patient preferences for key attributes associated with substantially different refractory OAB therapies. The findings indicate that, compared to Likert items, BWS is more sensitive and robust in capturing both positive and negative attributes driving treatment selection.

Valuing New HRQOL Measures: A DCE Application for Adverse Childhood Experiences and Maltreatment

Derek Brown, PhD, *Washington University in St Louis;* **Benjamin Craig, PhD,** *Moffitt Cancer Center;* **Xiangming Fang, PhD,** *China Agricultural University;* **Lisa Prosser, PhD,** *University of Michigan;* **Phaedra Corso, PhD,** *University of Georgia*

Purpose: To use DCE methods to value new HRQOL items (developed for capturing the burden of adverse childhood experiences and maltreatment) in terms of health state utilities for QALYs.

Methods: A set of seven (7) HRQOL items, each with 4 levels, and specific to adverse childhood experiences and maltreatment were developed following psychometric principles, focus groups, and pretesting. To use these items for future economic evaluation in terms of QALYs and health state utilities, a valuation study was created to estimate preference weights (i.e., a tariff or value set) for all possible 16,384 states. Valuation was implemented using a discrete choice experiment (DCE) fielded to a nationally representative online sample of n=1,864 U.S. adults (GfK's KnowledgePanel). The DCE tasks involved a two-part choice task. Respondents first selected which of two possible health states described by the 7 attributes was preferred. Next, respondents were asked whether they would prefer the alternative for 10 years, or the selected state for a reduced number of years (9.5, 9, 8, or 6 years: a form of time trade-off). Each respondent completed 9 of these two-part valuation tasks. (To derive age-specific health utilities, respondents completed 3 tasks for their current adult age, 3 tasks for a hypothetical teenage state, and 3 tasks for a hypothetical childhood state.) A near-orthogonal, balanced experimental design with 3 overlapping attributes was used, and respondents were randomized to 40 blocks of 9 tasks. A bivariate probit econometric model was used to account for the two-part choice task.

Results: Out of the 7 HRQOL items, the strongest preference (worst HRQOL states) was identified for "risky choices," followed by depression, and problems with healthy relationships. Physical pain & ability, anger, and illness/injury had smaller impacts on HRQOL. The smallest impact was found for "emotions out of control." Across the 4 levels (never, rarely, sometimes, or often), "rarely" incurred a statistically and economically meaningful impact only for risky choices (-0.15) and depression (-0.22). "Sometimes" implied a large drop (>.20) for risky choices and depression, but all items showed a substantial drop in health state utility (>0.40) for "often." Impacts were largest among teenage years, and larger (but less than teens) among younger children. Few respondents failed basic validity checks among the sample or rated the activities as difficult to complete.

Conclusions: This study demonstrates a novel and innovative approach to using DCEs to value health states and HRQOL measures. Key innovations include the two-part choice format, the overlapping DCE design, and the age-specific valuation period. These methods may be adapted in other contexts for valuing HRQOL where appropriate population or patient-specific value sets do not exist. On the subject of adverse childhood experiences and maltreatment, the estimated health state utility impacts may seem quite large. However, in another study sample, we found that the frequency of reporting "sometimes" or often" was quite low; based on the valuations here, the average combined marginal impact for severe maltreatment was -0.18 in health state utility among adults. "

The Effect of Framing of Death on Health State Values Obtained from Discrete Choice Experiments *Marcel Jonker, MSc, iBMG, Erasmus University Rotterdam;* ***Esther de Bekker-Grob, PhD,*** *MGZ, Erasmus MC;* ***Bas Donkers, PhD,*** *ESE, Erasmus University Rotterdam;* ***Elly Stolk, PhD,*** *iBMG, Erasmus University Rotterdam*

Purpose: DCE with duration as an attribute is considered a promising strategy for health state valuations. However, the implicit procedure for anchoring obtained values onto the full health-death scale conflicts with explicit decisions of health states such as obtained in Time Trade Off or DCE approaches with death included as an alternative-specific choice option. The purpose of this study is to test the hypothesis that those discrepancies occur because of different framings of 'death' in those tasks: implicit or explicit, immediate or postponed.

Methods: To test the impact of framing effects associated with 'death', an experiment with 4 distinctly different framings was conducted. These framings comprise both DCE approaches (i.e. DCEduration and DCEdeath) with and without the addition of lead time (LT) to the health profiles. A Bayesian efficient design consisting of 8 sets of 30 (matched pairwise) choice tasks was used, each set providing full identification of the EQ5D5L parameters for individual respondents. The design was jointly optimized for all framings, thereby keeping all aspects of the DCE design except for the framing constant. The choice task complexity for the respondent was kept to a minimum using level overlap and color coding. Respondents were randomly assigned to one of the 4 study arms. Mixed logit models were used to analyze the DCE data, and the resulting estimates of the utility decrements associated with the severity levels within each dimension were compared between the 4 arms to establish the impact of the framing effects.

Results: The DCE was administered to a Dutch nationally representative sample of 1200 respondents equally divided over the 4 study arms. The estimation results revealed substantial framing effects. While the DCE death approach classified just 8% of the health states as worse than death, much higher percentages were found in the other arms: 28% (duration), 57% (LT-death) and 81% (LT-duration). Relative distances between health states on the latent scale were not affected by adding LT, but anchoring on death altered the values. We observed less dispersion for mild to moderate states, and a more stretched distribution for severe states.

Conclusions: Estimation results were substantially altered by the framing of death as explicit or implicit, and immediate or postponed. These framing effects may help to explain the commonly observed discrepancies between values derived using Time Trade Off and the popular DCE duration approach. While one may argue against the use of a death alternative in DCE tasks for health state valuation on basis of theoretical and statistical considerations, it would seem to be an essential component for those who aim to reconcile DCE and TTO results.

Framing of Attribute's Levels: Influence on the Interpretation of Outcomes from a BWS Experiment.

Marieke Weernink, MSc; Karin Groothuis-Oudshoorn, PhD; Maarten IJzerman, PhD; and *Janine van Til, PhD,* University of Twente

Purpose: In a previous study, a Best-Worst Scaling case 2 (BWS) was conducted to elicit treatment preferences for symptom control, side-effects, and process characteristics of various treatments prescribed in Parkinson's Disease. It is known that people, especially elderly, have trouble interpreting risks. Therefore it was decided to decrease the cognitive load by using a qualitative operationalization of the levels of symptoms and side-effects. The disadvantage of this strategy was that patients might have different perceptions of risks occurring "seldom to never, sometimes or often". Our objective was to study whether this operationalization influenced the interpretation and valuation of outcomes of our BWS experiment.

Methods: A post-questionnaire was distributed among 30 patients who participated in the BWS-study. Patients were asked to qualitatively state their perception of the extent to which they suffered from the symptoms and side-effects of treatment: tremors, posture and balance problems, slowness of movement, dizziness, fatigue, and dyskinesia. Subsequently, they were asked to quantify this extent, as well as their (quantified) perception of the other two levels. In analysis, patients were grouped based on their experienced suffering state per attribute, medians were estimated, and a perceived duration range for each attribute was estimated. Beside, an actual duration range for each attribute was estimated based on the data of the experienced states. Based on these results, subgroup analyses of our original BWS data were performed to study whether the experienced amount of suffering from an attribute influenced attribute importance.

Results: The 30 patients indicated a wide range of perceived severity of symptoms and side-effects of treatment. For each symptom or side-effect, we included patients that perceived their extent as "seldom", "sometimes" or "often". Interestingly, when asked to quantify the extent of suffering, the perceived impact of suffering from a symptom (within respondent) is smaller than the actual range (between respondents). The perceived burden of a symptom or side-effect is underestimated most by patients that seldom to never suffer from an attribute with a factor ranging from 1.5 - 8 for the different attributes. This underestimation was also present for patients who sometimes or often suffered from an attribute, but was much smaller: ranging from 1.1 - 1.6. BWS subgroup analysis on the original data showed that patients who seldom suffer from a symptom assigned a lower importance to that symptom compared to patients who often suffer from it, indicating that this underestimation is present in the original data as well. These differences ranged from 2% for posture and balance problems to 11% for the suffering of dyskinesia (between seldom to never and often suffer from).

Conclusions: The qualitative operationalization of attribute levels resulted in different interpretations between the subjects and influenced BWS results. Patients who only have minor complaints from a symptom or side-effect, seem to underestimate the actual burden of having major complaints. As a result,

patients assigned a lower attribute importance to attributes from which they only seldom to never suffer from and a higher attribute importance to attributes from which they often suffer from.

1:00-2:30 PM   **Session 3**

Attribute Non-Attendance and Time Pressure in Discrete Choice Experiments: An Eye-Tracking Study

Benjamin Cooper, City University London; <u>Kate, Mandeville, BSc, MBBS, MSc,</u> London School of Hygiene & Tropical Medicine; Kielan Yarrow, BSc, MSc, PhD, City University London

Purpose: To investigate attribute non-attendance in a discrete choice experiment under different time conditions using eye-tracking

Methods: The design of the discrete choice experiment was based on an existing study examining preferences for primary care appointments. 48 participants completed 25 choice tasks (24 novel and 1 repeated) whilst having their eye movements recorded by an eye-tracking camera. Participants were divided into two groups: self-paced or computer-paced. The first group were able to progress at their own rate, whereas the second were forced to wait 25 seconds before being able to make a choice and move on to the next task. The outcome measure was time spent looking at different attributes, particularly the attributes on which the participants spent the most and least attention.

Results: There was a significant difference in time spent looking at different attributes between the two groups. The computer-paced group spent significantly longer looking at their most and least attended attributes in all choice tasks compared to the self-paced group. The self-paced group spent significantly less time looking at their most and least attended attributes as the experiment progressed, with some participants spending no time on certain attributes. In contrast, computer-paced participants spent approximately the same amount of time looking at their most attended attribute throughout the experiment. However, they also spent less time looking at their least attended attribute as the experiment progressed.

Conclusions: Attribute non-attendance was confirmed through the use of eye-tracking, suggesting the use of decision heuristics. The use of a time condition may discourage the use of such heuristics in discrete choice experiments.

Respondent Cognition in Health Preference Research

Benjamin M. Craig, PhD, Moffitt Cancer Center and University of South Florida; <u>Shannon K. Runge, MA,</u> University of South Florida and Moffitt Cancer Center

Purpose: Recent studies have emphasized the need to better understand the cognitive burden of discrete choice experiments (e.g., eye-tracking) to improve their design. Yet, no health preference study has directly incorporated a cognitive assessment, such as a memory test, to assess (1) whether those with lower cognition respond more randomly or (2) whether those who experience greater cognitive decline respond more randomly.

Methods: As a measure of episodic memory, the immediate and delayed word recall tasks from the US Health and Retirement Study (HRS) were adapted for inclusion in the Women's Health Valuation (WHV) study. Between the 2 recall tasks, the survey asked respondents (women, ages 40 to 69) to complete 30 paired comparisons, each with two menopausal symptoms described using the Patient-Reported Outcomes Version of the Common Terminology Criteria for Adverse Events (PRO-CTCAE) and the question, "Which do you prefer?" As a measure of randomness, the root likelihood was introduced as the geometric mean of a response pattern that can be rescaled at the respondent-level, such that 0 represents random response and 1 represents always giving the majority response. This study tests the association between a respondent's self-reported memory, word recall and her root likelihood.

Results: Among the 3397 respondents, self-reported memory ranged from Excellent to Poor (Excellent/VeryGood/Good: 88%; Fair/Poor: 12%), immediate word recall ranged from 0 and 10 (mean 7.24; IQR 6-8), the difference between immediate and delayed recall ranged from 0 to 9 (mean 1.15; IQR 0-2), and

the root likelihood ranged between -0.36 and 1 (mean 0.66; IQR 0.56-0.82). Self-rated memory is mildly associated with immediate recall (Excellent/VeryGood/Good: 7.34; Fair/Poor: 6.5; p-value <0.01) and root likelihood (Excellent/VeryGood/Good: 0.67; Fair/Poor: 0.65; p-value 0.038). For the immediate recall, the root likelihood linearly increased for each word recalled until 7 words (i.e., median) then flattens. The difference in recall was not associated with the root likelihood.

Conclusions: The results suggest that worse memory is associated with greater randomness in DCE response. One possible explanation is that persons with "poor" memories have eccentric preferences; however, the more plausible interpretation is that DCEs are cognitively challenging. More research is needed to better recognize design issues among those cognitively challenged (particularly among older adults) and to identify strategies that ameliorate the cognitive burden of health preference studies.

Mitigating Hypothetical Bias in Stated Preference Discrete Choice Experiments

Dean Regier, PhD, BC Cancer Agency, Canadian Centre for Applied Research in Cancer Control, and University of British Columbia; *Trevor Meier, BA,* Story Spark Media

Purpose: Stated preference methods are used to estimate the value of non-marketed commodities. Because of the hypothetical nature of the tasks, respondents' stated preferences may not coincide with the choices they would make in "real-world" decision contexts. This is called "hypothetical bias". Recent research suggests that hypothetical bias can be attributed to a lack of "informational base" or time to think. With a view to ameliorate the effects of hypothetical bias, the objective of this study was to create a video introduction to a stated preference discrete choice experiment survey (DCE) and evaluate its influence on willingness to pay (WTP).

Method: The DCE study elicited public values for the return of incidental findings obtained from next generation sequencing technology. Focus groups were used to identify 5 attributes and pre-test interviews were conducted to ensure subjects understood the choice tasks. A video introduction was created using computer animation with a voice layover describing the DCE attributes and emphasizing the importance of considering all attributes, including cost (i.e. cheap talk). Critical comments regarding the video introduction were provided through 10 in-person interviews with the public. Respondents to the DCE were randomized to a video-introduction DCE survey (followed by a text description) or a standard text-only introduction. Split sample analysis (text-only sample versus video introduction sample) was undertaken using the mixed logit model. Confidence intervals for WTP estimates for each sample were generated using the Krinsky-Robb method.

Results: A total of 1200 respondents completed either questionnaire. Respondent characteristics between the two questionnaire versions were non-statistically significantly different from each other. In both samples the part-worth utility values were statistically significantly different from no effect on indirect utility, and measures of model fit using the Pseudo R2 and the Bayesian Information Criterion were similar. We examined several WTP scenarios using the compensating variation formula. In each of the examined scenarios, the sample completing the video survey introduction DCE had lower mean WTP estimates compared to the text-only introduction. The p-value examining WTP differences between the populations ranged from 0.095 to 0.11.

Conclusion: We found that improving respondents information base through a video introduction to a DCE survey may mitigate hypothetical bias compared to text only DCEs.

Discount Rate Assessment among Adults Experiencing Dyspnea from Common Primary Care Diseases

Irene D. Fischer, MPH, and **Walton Sumner, MD,** Washington University; **Michael D. Hagen, MD,** American Board of Family Medicine

Purpose: Discount rates (DR) for health are an important patient preference for cost-effectiveness analyses, but they are seldom directly assessed. Previous research implies that DR can be negative. The standard guidance for cost-effectiveness analyses is to use a DR of 3% (0.03).

Methods: We conducted an experiment using 3 different methods to estimate individuals' DR: 1) Forced choice methods presenting several combinations of delayed relief and transient quality of life

improvements, involving (a) rationing pain relief medication or (b) temporary blindness cures. In context, each combination suggested a range of DR; 2) Sequential binary choices that re-used the same choices presented in (1), but also permitted subjects to declare indifference between two choices, and permitted combinations of responses that are inconsistent with a fixed DR; 3) Direct assessment of DR using intersecting survival curves with a fixed quality of life. Subjects were recruited for the computer-guided interview via a volunteer database available through the university. Beyond the basic remuneration, they were offered financial bonus incentives for accurately interpreting survival curves and for being internally consistent with their choices. A numeracy assessment was included at the beginning of the interview. Over time different versions of the interview were used, including different combinations of discount rate assessments.

Results: 1a) Choices for rationing pain relief medication over 90 days indicated DR very close to zero over that time frame. 1b) Choices for temporary blindness cures over a decade resulted in 5% with DR <-0.05, 21% with DR between -0.05 and 0.05, 19% with DR between 0.05 and 0.15, and 56% with DR >0.15. 2) Sequences of binary choices yielded 0.6% results consistent with DR<-0.05, 14% with DR between -0.05 & 0.05, 9% with DR between 0.05 & 0.15, 38% with DR >0.15, and 38% inconsistent with a fixed DR. The distribution of results consistent with a fixed DR did not differ significantly from the distribution obtained in (1b). 3) Direct assessments yielded mean scores of -0.01 to 0.40, with males (-0.006 vs. 0.095 for monoblind DR, $p<0.05$; -0.06 vs. 0.15 for deaf blind DR, $p<0.05$) and respondents with a high level of internal consistency (-0.02 vs. 0.09 for monoblind DR, $p<0.05$) being more likely to have discount rates close to or below zero.

Conclusions: It is difficult to elicit individual DR estimates. Consistency checks can detect respondents whose results suggest major axiomatic departures or misunderstanding of the elicitation task. Over periods less than one year, DR are likely to be zero for symptom relief. Over ten years, several results suggest small negative DR for some subgroups of respondents, and positive DR much greater than 0.03 for many others. These results are challenging to incorporate in conventional cost-effectiveness analyses.

2:45-3:45 PM   Session 4 ─────────────────────────

Conjoint Analysis: A Tool for Understanding Patients Decisions for Invasive Treatments

Leslie Wilson, PhD; Tracy Kuo Lin, PhD; Lindsay A. Hampson, MD; Anna Oh, MSN, MPH, RN; Jie Ting, PhD; and Benjamin N. Breyer, MD, University of California, San Francisco

Purpose: Urinary stricture affects 1/1,000 men >= 65. Understanding patient preferences for treatment decisions places emphasis on patient-centered outcomes and shared decision-making, especially when clinical outcomes are unclear as for urinary stricture. Conjoint analysis is a method for understanding what risks patients are willing to take to gain treatment benefits. The study objective is to evaluate how patients, with urethral stricture, value characteristics of different treatments (open surgical reconstruction versus endoscopic incision).

Methods: We adopted a random, full-profile choice-based conjoint (CBC) design to maximize responses on urethral stricture treatment questions. CBC is an increasingly utilized methodology to systematically elicits risk-benefit trade-offs and quantifies the relative preferences and risk acceptance for attributes of medical interventions. We constructed our CBC survey to mimic the decision-making process of patients with urethral stricture disease choosing a treatment procedure. The hypothetical scenarios presented were based on realistically possible combinations of attributes. Each patient was presented with choices across six attributes: copayment amount, invasive vs non-invasive procedure, success rate, number of future procedures, post-procedure catheter duration, and recovery time. We employed a mixed-effects logistic regression model to analyze the CBC preferences.

Results: Among the six attributes this study examined, the success rate was by far the most important attribute. Relative to an 85% success rate: common after surgery: (OR = 1) a 25% success rate (OR =.037, $p<.01$) decreased patient preference by approximately 27 times, while a 50% success rate (OR=.18, $p<.01$) reduced patient preference by approximately 5 times. In addition, patients demonstrated a strong aversion to time with a urinary catheter, more common with surgery. Catheter duration for 1 week or less (OR =

.665, p<.01) reduced patient preference by about 1.5 times when compared to requiring no catheter. Similarly, patients showed that their preferences decreased as the number of procedures increased. Patients' negative preference for copayment amount was not statistically significant until the copayment amount was $1000. Relatively to no copayment, a $1,000 copayment (OR=.699, p<.01) decreased patient preference by approximately 1.5 times, while $10,000 (OR =.228, p<.01) decreased patient preference by about 4.4 times. Patients did not have a statistically significant preference for recovery time which can be longer for surgery, indicating that patients placed more importance on other attributes included in the study. Relative to an invasive procedure, patient showed a clear negative preference for a non-invasive procedure (OR=.815, p<.01). However, the difference in patient preference was not as large as when compared to other attributes, suggesting that patients value other attributes over procedure type itself. **Conclusions:** The findings show that patients value the characteristics and outcomes of procedures more than the invasiveness of the procedure. With the exception of the characteristic of catheter duration, patients prefer characteristics associated with open reconstruction; however, the preference for an invasive procedure itself is relatively small. The results suggest that patients may not be clear about the characteristics of different procedures when presented as surgical or non-surgical. It is crucial to ensure patients understand all aspects of treatments in the shared decision-making process.

Open Discussion
Derek Brown, PhD, MA, BA, *Washington University in St. Louis*

Business Agenda

1. Opening and agenda: Derek Brown, Meeting Chair

2. Financial report

3. Membership report

4. Announcements of future meetings

5. Discussion on IAHPR logo

6. Discussion on book and manuscript opportunities

7. Collaborations with software vendors

8. Discussion on sustainability

9. Closing

STUDENTS AND THEIR ADVISORS

STUDENTS	ADVISORS
Johns Hopkins Bloomberg School of Public Health, Baltimore, Maryland, USA	
Ilene Hollin, MPH	John Bridges, PhD*
University of South Florida, Tampa, Florida, USA	
John Hartman, MA	Benjamin M. Craig, PhD*
Shannon Runge, MA	Benjamin M. Craig, PhD*
London School of Hygiene and Tropical Medicine, United Kingdom	
Kate Mandeville, MBBS, MSc, BSc	Kara Hanson, SD and Mylene Lagarde, PhD, MSc, MA
University of Maryland, Baltimore, Maryland, USA	
Max Masnick, BA	Anthony Harris, MD, MPH
University of Cape Town, Cape Town, South Africa	
Bernard Njau	Jan Ostermann, PhD*
University of Minnesota, Minneapolis, Minnesota, USA	
Sarah Lillie, PhD, MPH	Melissa Partin, PhD
Instituto Nacional de Cardiología, Rio de Janeiro, Brazil	
Andréa Monteiro	Marisa de Silva Santos, MD, PhD
University of Twente, Enschede, The Netherlands	
Marieke Weernink	Janine van Til, PhD
University of California, San Francisco, San Francisco, California, USA	
Tracy Lin, PhD	Leslie Wilson, PhD*

*IAHPR Member; alphabetized by advisor.

Current IAHPR Members

Esther de Bekker-Grob*

Kathleen Beusterien*

Meenakshi Bewtra*

Abhijeet Bhanegaonkar

John Bridges

Derek Brown*

Margaret Byrne

Benjamin M. Craig*

Terry Flynn*

Liana Fraenkel*

Juan Marcos González

Karin Groothuis-Oudshoorn

A. Brett Hauber

Sarah Hawley*

Joel Hay

Kirsten Howard

Joel Huber

F. Reed Johnson*

Christine Kistler*

Emily Lancsar

Deborah Marshall

Ateesha Mohamed*

Axel Mühlbacher*

Mark Oppe

Jan Ostermann*

Christine Poulos

Lisa Prosser*

Shelby Reed*

Dean Regier*

Mandy Ryan

Christopher Saigal*

Elly Stolk

Jamie Studts

Jennifer Whitty

Leslie Wilson

* Attending

Upcoming IAHPR Meetings

4th Meeting of the International Academy of Health Preference Research
Wednesday, 13 July 2016, 8:00-17:30, chaired by Mandy Ryan and Elly Stolk
Nord Event Panoramadeck, Emperio Tower, Dammtorwall 15, Hamburg, Germany

5th Meeting of the International Academy of Health Preference Research
Friday, 2 September 2016, 8:00-17:30, chaired by Kirsten Howard and Mark Oppe
hosted by the Center for Health Services and Policy Research (CHSPR) at the Saw Swee Hock
School of Public Health (SSHSPH) and the National University Hospital System
of Singapore (NUHS), Tahir Foundation Building, National University of Singapore, Singapore

6th Meeting of the International Academy of Health Preference Research (Tentative)
Friday, 7 July 2017, chaired by Juan Marcos González and Joel W. Hay
TBD, Boston, Massachusetts, USA

7th Meeting of the International Academy of Health Preference Research (Tentative)
November 2017, chaired by Karin Groothuis-Oudshoorn and Terry Flynn
TBD, Glasgow, Scotland, UK

8th Meeting of the International Academy of Health Preference Research (Tentative)
September 2018, chaired by Brendan Mulhern and Jennifer Whitty
TBN, Australia

9th Meeting of the International Academy of Health Preference Research (Tentative)
13 October 2018, chaired by TBN
Fairmont The Queen Elizabeth, Montréal, Québec, Canada

10th Meeting of the International Academy of Health Preference Research (Tentative)
September 2019, chaired by TBN
TBN, Europe