

OBJECTIVES: To explore how participants evaluate and complete the choice tasks in Discrete Choice Experiments (DCE), with special attention to the impact of educational level and health literacy. **METHODS:** Two existing DCE questionnaires on rotavirus vaccination and prostate cancer screening served as a case for the current study. In total, 70 participants were sampled based on educational level (35 per case study). During structured interviews, participants completed five choice tasks aloud. Interviewers monitored how participants read the choice tasks, how they interpreted the included risk attributes and what decision strategy they used to make their decision and if the monotonicity and continuity axioms hold. **RESULTS:** The majority of the participants read all the attributes within each choice task. Nearly all participants chose the scenario with the optimal attribute levels (monotonicity axiom). In accordance with the continuity axiom, most participants mentioned three or more attributes when motivating their decisions. Overall, higher educated and literate participants more often included three or more attributes when motivating their decision and used trading between attributes more often as a decision strategy. **CONCLUSIONS:** The majority of the participants complete a DCE as presumed by its underlying methodology. However, the assumptions did not hold for a subset of lower educated and less literate participants. Based on participants' age, educational level and health literacy additional measures should be undertaken to enhance participants' understanding of the attributes, the attribute levels and the choice tasks in a DCE.

PRM140

THE MEASUREMENT OF UTILITIES IN ASTHMA PATIENTS: A PRELIMINARY STUDY

Yong YV, Shafie AA

Universiti Sains Malaysia, Penang, Malaysia

OBJECTIVES: To assess the feasibility of a computer-based Standard Gamble (SG) visual prop whilst measuring utilities of different asthma health states at the same time. **METHODS:** Twenty adult asthma patients literate in either Malay or English language were conveniently sampled from a public hospital in Penang, Malaysia. They were interviewed by two trained interviewers using a bilingual script. Each patient was requested to value the given health states using Visual Analogue Scale (VAS) prior to SG exercise. There were three chronic health states (C1-C3) for 10 years, three temporary states (T1-T3) for 3 months, and two anchor states (healthy and dead). During the SG exercise, the visual prop was fully operated by the interviewers. The probability of being in a worse state was changed in a 'ping-pong' fashion until the indifference point was reached. **RESULTS:** All patients understood the SG exercise and rated SG easier than VAS. Around 85% (n=17) completed SG within 30 minutes. There was 90% (n=18) who ranked T3 as the worst temporary health state during VAS. Two patients provided logical inconsistency data in SG. The preferences by SG were higher than VAS. Preferences were also higher in temporary states measured by chained SG than other states by conventional SG. The mean utilities for C1=0.56 (SD 0.38), C2=0.47 (SD 0.33), C3=0.53 (SD 0.38), T1=0.65 (SD 0.31), T2=0.53 (SD 0.35), and T3=0.38 (SD 0.38). **CONCLUSIONS:** The SG methods including the props are feasible for utilities measurement in asthma, based on the agreements achieved with other studies on the pattern of utilities measured in this preliminary study.

PRM141

DISCRETE-CHOICE EXPERIMENT VERSUS RATING SCALE EXERCISE TO EVALUATE THE RELATIVE IMPORTANCE OF ATTRIBUTES: A STUDY OF THE MAASTRICHT ISPOR STUDENT CHAPTER

Wijnen BFM¹, van der Putten IM¹, Groothuis S¹, de Kinderen RJA¹, Noben CYG¹, Paulus ATG¹, Ramaekers BLT², Vogel GCWM², Hilgsmann M¹¹Maastricht University, Maastricht, The Netherlands, ²Maastricht University Medical Center, Maastricht, The Netherlands

OBJECTIVES: Eliciting preferences has become increasingly important in health care. Several methods are available to evaluate the relative importance of different aspects of health and health care. In this study, we aim to examine the difference between a discrete-choice experiment (DCE) and a rating scale exercise (RSE) to determine the most important attributes of undergraduate students when selecting a study specialization. **METHODS:** First-year health sciences students were asked to complete a questionnaire that included a DCE and a RSE. Six attributes were identified in focus groups: "possible acquainted masters", "job opportunity", "scope of specialization", "quality of education", "hours self-study" and "personal interest". Fourteen unlabeled choice tasks were constructed using a statistically efficient design and a mixed multinomial logistic regression analysis was used for data analysis. In the RSE, attributes were rated on perceived importance using a 7-point Likert scale. Two versions of the questionnaire were distributed in which the RSE was put before and after the DCE. **RESULTS:** A total of 254 students filled out the questionnaire. In the DCE, three attributes were statistically significant of which "personal interest" was the most important attribute followed by "job opportunity" and "quality of education". In the RSE, all attributes except "hours of self-study" were rated 4 or higher. The RSE scores of the attributes with a relatively low importance in the DCE were significantly lower in the questionnaire version that started with DCE than in the other version. Results of the DCE did not significantly differ between the two questionnaire versions. **CONCLUSIONS:** The DCE had a differentiating effect on the relative importance of attributes whereas in the RSE attributes were rated more equally and were, except for one, all considered important. Forcing respondents to make first trade-offs between attributes (using a DCE) leads to lower RSE scores for less important attributes afterwards.

PRM142

THE DEVELOPMENT AND PRELIMINARY VALIDATION OF THE MANCHESTER SLEEP SYMPTOMS INDEX (MSSI) FOR PEOPLE WITH CHRONIC OBSTRUCTIVE PULMONARY DISEASE (COPD)

Garrow AP¹, Yorke J¹, Khan N², Tyson S¹, Singh D³, Vestbo J⁴¹Manchester, Manchester, UK, ²The Medicines Evaluation Unit, Manchester, UK,³University of Manchester, Medicines Evaluation Unit, Manchester, UK, ⁴The University of

Manchester, Manchester, UK

OBJECTIVES: In COPD, disturbed sleep is related to exacerbation frequency, poor quality of life and early mortality. We developed the Manchester Sleep Symptoms Index (MSSI) to assess night-time symptoms and disturbed sleep in COPD. **METHODS:** Identification of potential items was guided by interviews and focus groups involving COPD patients and age-matched controls. Hierarchical methods and Rasch analysis informed item deletion and development of a unidimensional scale. Internal consistency and test-retest reliability were assessed. Concurrent validity was examined using Pearson's correlation with the St George's Respiratory Questionnaire (SGRQ), modified MRC Dyspnoea Scale, FACIT fatigue scale, MOS Sleep Scale, HADS Anxiety and Depression scores. **RESULTS:** Qualitative data from 36 patients with COPD and 10 age-matched controls informed an initial list of 22 items. The cross-sectional study included 203 COPD patients (GOLD: I: 14% II: 41% III: 25% IV: 7%; male: 63%, mean age: 64.7; SD: 7.5 years) and 50 non-COPD controls. 12 items were removed during hierarchical methods and a further two following Rasch analysis. The final MSSI contains 8 sleep-related items that are specific to COPD patients: breathlessness, chest tightness, cough and sputum production. The index has good internal consistency (Cronbach's alpha 0.87), test-retest repeatability (intra-class coefficient 0.77) and validity. Total MSSI scores significantly correlated with the SGRQ (r=0.64); MRC Dyspnoea scale (r=0.46), FACIT-F (r=-0.61); MOS problems index 2: (r=0.62); MOS Sleep adequacy (r=0.40); MOS Sleep disturbance (r=0.53) HADS anxiety (r=0.54) and depression (0.48). There was good overall fit to the Rasch model (Chi-squared: 29.2 df: 16 p=0.03) and distribution of item scores. **CONCLUSIONS:** The MSSI is a reliable, valid, uni-dimensional self-reported outcome measure of sleep and night-time symptoms for people with COPD. It is simple and quick to use making it suitable for research and practice. Further work is needed to determine the minimal clinical important difference and cross cultural validity.

PRM143

PSYCHOMETRIC EVALUATION OF THE PATIENT'S KNEE IMPLANT PERFORMANCE (PKIP) QUESTIONNAIRE FOR THE ASSESSMENT OF PRIMARY TOTAL KNEE ARTHROPLASTY

Coles TM¹, Dwyer KA², Mordin M³, Williams V¹, Clatworthy M⁴, Yates P⁵, Hamilton W⁶¹RTI Health Solutions, Research Triangle Park, NC, USA, ²DePuy Orthopaedics, Inc, Warsaw, IN, USA, ³RTI Health Solutions, Ann Arbor, MI, USA, ⁴Ascot Hospital, Auckland, New Zealand, ⁵Murdoch Orthopaedic Clinic, Murdoch, New Zealand, ⁶Anderson Orthopaedic Research Institute, Alexandria, VA, USA

OBJECTIVES: The objective of this study was to evaluate the psychometric properties of a new patient-reported measure of knee implant functional performance associated with physical activities prior to and following primary total knee arthroplasty (TKA). The Patient's Knee Implant Performance questionnaire (PKIP) was developed to assess factors that lead to patient dissatisfaction and describe unmet needs in knee functional performance. **METHODS:** The psychometric analysis sample (n=764) was based on a multicenter, prospective, noncomparative longitudinal study of patients with osteoarthritis undergoing TKA at 22 international sites. The PKIP and additional patient-reported outcomes and clinical measures were collected preoperatively, postoperatively at less than 1 year, at a minimum of 1 year, and at 2 years. The PKIP structure and its reliability, construct validity, discriminating ability, and responsiveness were assessed. **RESULTS:** Based on inter-item correlations, factor analyses, and results of previous qualitative research, the PKIP was scored as four subscales (Stability, Confidence, Satisfaction, and Activity Modification) and an Overall PKIP score. The Overall PKIP score met reliability standards (internal consistency: alpha =0.78 at minimum 1 year; test-retest: intraclass correlation coefficient =0.77). Correlations between the PKIP and other available measures provided evidence of construct validity. For example, the PKIP correlated 0.19 and 0.50 with the American Knee Society Score preoperatively and at less than 1 year, respectively, and correlated 0.69 and 0.77, with the Knee Injury and Osteoarthritis Outcome Score Quality of Life subscale. The PKIP was capable of discriminating between groups of patients with better or worse knee functioning as defined by clinician-rated measures; hypothesis tests were in the predicted direction and mostly statistically significant. The effect size for the Overall PKIP score was 2.38, indicating that the PKIP was highly responsive. **CONCLUSIONS:** The reliability, validity, and responsiveness of the PKIP support its use among patients undergoing primary TKA.

PRM144

VALIDITY AND RESPONSIVENESS OF THE BRISTOL RHEUMATOID ARTHRITIS FATIGUE MULTIDIMENSIONAL QUESTIONNAIRE (BRAFM-DQ) IN A RANDOMIZED CONTROLLED CLINICAL TRIAL

Kirwan J¹, Coteur G², Dures E¹, Nicklin J¹, Bryson J³, Hewlett S¹¹Bristol Royal Infirmary, Bristol, UK, ²UCB Pharma, Brussels, Belgium, ³UCB Pharma, Raleigh, NC, USA

OBJECTIVES: To evaluate the validity of the BRAFM-DQ in a new group of patients in a clinical trial setting, to confirm its internal factor (domain) structure and to document its sensitivity to change. **METHODS:** Pooled data from a randomized controlled trial (NCT01242488) in patients with moderate to severe RA were collected at baseline (BL), Wk10 and Wk12. Spearman's correlation coefficients, Bland-Altman plots and confirmatory factor analysis tested construct validity, reproducibility and internal factor structure of the BRAFM-DQ. Responsiveness was assessed amongst clinical responders at Wk12 by effect sizes of changes from BL in BRAFM-DQ. **RESULTS:** There were 219 patients (mean age: 55.5 years; disease duration: 12.6 years; BL DAS28: 5.77). The proportion of missing item answers was very low (0-3%). BRAFM-DQ scores correlated with patient global, pain and HAQ at BL (r=0.49, 0.46 and 0.58) and at Wk12 (r=0.63, 0.65 and 0.64). Changes in the Physical and Living domains were more closely related to changes in patient and physician global scores and DAS scores than did changes in Cognition and Emotional domains. Reproducibility was high (r>0.87 for all total and domain scores; narrow Bland-Altman limits of agreement), as was internal consistency (Cronbach's alpha: 0.97 for total scores; >0.82 for each domain at BL). The Butler comparative fit index (CFI; 0.92) indicated that the established structure within the BRAFM-DQ accounts well for data variation. Effect sizes for BRAFM-DQ in clinical responders at Wk12 were very high