mon use across and between conditions. **METHODS:** Subjects were recruited by web posting and telephone screening. Those self-reporting active treatment for one of four conditions (Migraine, Low Back Pain (LBP), Osteo-Arthritis (OA), Rheumatoid Arthritis (RA)) were scheduled for in-person interviews using card sort exercises with 54 different pain descriptors to identify those each subject commonly used to describe the pain associated with their condition. **RESULTS:** Subjects ranged between 29 and 63 years (mean age of 45). The majority (71%) was female, 60% were working full or part time, and 51% were Caucasian. Pain descriptors were divided into three groups based on percent of subject endorsement; 70-100% for high use 45-69% for moderate use, and 18-44% for low use. Across all four conditions, the most used pain descriptors were SHARP, ACHING, THROBBING, and HURTING. Moderate use was shown across all conditions for terms like RADIATING, SPREADING, STINGING, and JOLTING. Lower percentages of study subjects used descriptors with more specific pain characteristics, including temperature (HOT, SEARING, BURNING), neuropathic characteristics (NUMB, PRICKLING) and qualities associated with acute pain (CUTTING, TEARING). Migraine subjects tended to use descriptors like POUNDING (83%), SPLITTING (88%) and PULSATING (77%). LBP subjects used STABBING (77%), PINCHING (84%), and SHOOTING (74%). OA and RA patients tended to be more similar with CONSTANT (46%/73%), TENDER (55%/64%) and SORE (72/73%). Additional descriptors unique to RA included TIGHT (73%) and CRAMPING (64%). **CONCLUSIONS:** Because descriptors of pain used by patients across these four different conditions showed use of similar language as well as expressions that were unique to their condition, the assessment of condition specific pain should be considered when planning to use pain as a study endpoint.

PRM92

EXPLORATORY MAPPING OF THE LUPUSQOL TO THE SF-6D

Meacock R[1], Harrison M[1], Mcelhone K[2], Isenberg D[3], Ferenkeh-Koroma A[4], Ahmad Y[5], Bruce IN[1], Shelmerdine J[6], Gordon C[7], Griffiths B[8], Maddison P[9], Akil M[10], Abbott J[11], Teh LS[2]

[1]*University of Manchester, Manchester, UK,* [2]*Royal Blackburn Hospital, Blackburn, UK,* [3]*University College London Hospitals, London, UK,* [4]*University College London, London, UK,* [5]*Betsi Cadwaldr University Health Board, Llandudno, UK,* [6]*Macclesfield District General Hospital, Macclesfield, UK,* [7]*University of Birmingham, Birmingham, UK,* [8]*Freeman Hospital, Newcastle, UK,* [9]*Bangor University, Bangor, UK,* [10]*Royal Hallamshire Hospital, Sheffield, UK,* [11]*University of Central Lancashire, Preston, UK*

OBJECTIVES: To derive a mapping algorithm to estimate scores (utility values) for the preference-based SF-6D measures from the non-preference-based disease-specific LupusQoL. **METHODS:** A total of 282 systemic lupus erythematosus (SLE) patients completed the LupusQoL and SF-6D at the same assessment. Models of the relationship between them were estimated using OLS regression. The SF-6D utility score was modelled using total scores on the 8 LupusQoL domains, employing a backward inclusion procedure. Model performance was judged using the root mean squared error (RMSE) and range of predicted values. **RESULTS:** The mean (SD) age of the sample was 45 (13.4) years and the mean (SD) SF-6D score was 0.61 (0.13). The mean scores for the LupusQoL domains ranged from 52.5 (Fatigue) to 73.5 (Body Image). Four of the eight LupusQoL domains were selected for inclusion in the final model (Physical Health, Pain, Emotional Health, Fatigue) because these domains were measured in both instruments. The root mean square error (RMSE) for the mapping function was 0.0701, lower than that reported for many published mapping functions. The overall model fit was good (R^2=0.7155), although some under prediction at the upper end of the SF-6D was observed. **CONCLUSIONS:** There appears to be a strong relationship between the LupusQoL and SF-6D. Prediction errors are lower than for many published mapping functions, signifying that the mapping algorithm developed here provides a methodology for predicting SF-6D utility values from LupusQoL data. Potentially this could reduce patient burden if all of the necessary information can be obtain from administering the LupusQoL alone. However, the omission of disease-specific LupusQoL domains (intimate relationships, body image, burden to others, planning) from the final model, raises concerns that the specificity for SLE may be lost in this algorithm. Further out of sample testing will be useful to confirm the performance of this algorithm.

PRM93

INFANTILE HEMANGIOMA ON FAMILIES,CREATION OF A SPECIFIC BURDEN QUESTIONNAIRE

Boccara O[1], Meni C[1], Labreze C[2], Bodemer C[3], Voisard JJ[4], Taieb C[5]

[1]*Necker Hospital, Paris, France,* [2]*Hospital Pellegrin, Bordeaux, France,* [3]*René Descartes University, Paris, France,* [4]*PFD, Lavaur, France,* [5]*CREES PFSA, Boulogne, France*

OBJECTIVES: Infantile hemangioma (IH) is the most common form of benign vascular tumor in children, with an estimated incidence of between 3 and 10% of Caucasian children. The aim of our study is to develop a specific questionnaire for assessing the burden on families of children with HI. **METHODS:** A "Hemangioma Family Burden" questionnaire (HFB) consisting of 22 items. It was distributed accompanied by SF12 and PGWBI, to obtain internal and external validation. **RESULTS:** Fifty-eight evaluable questionnaires were returned, completed by either the mother (63.27%) or the father (36.7%). At the time of the survey, the average age of the child with IH was 9.34±4.75months. Internal validity was measured by Cronbach's-alpha (0.95), reflecting a good homogeneity of the 22 items.While the score of the physical component does not vary with the expressed extent of the hemangioma (p=0.2931), the burden scores of the SF12 mental component and the PGWBI "wellness" component increase with "severity", as perceived by the parents and expressed in terms of extent. Hence, the HFB score is correlated with these 2 components, thus confirming external validity.The mean score calculated from the HFB is 23.42±19.93. The score increases with the "severity score" of the parents. In fact, a statistically significant difference is observed between the 3 severity groups: 5.28±6.8 for those reporting the smallest extent to 41.0±18.71 for those reporting the greatest extent, and 27.7±16.96 for a moderate extent. This confirms the sensitivity of the HFB. **CONCLUSIONS:** Internal and external validity were confirmed. The HFB is correlated with the extent felt by parents, a feeling deemed relevant because it is often the cause of consultation and demand for treatment. We now have an easy-to-use, validated IH tool for assessing the disability caused. Following cultural and linguistic validation, the HFB is now available in US English, Spanish, German and Italian.

PRM94

A NEW STEP IN THE USE OF COMPUTER ADAPTIVE TESTS FOR MEASURING QUALITY OF LIFE. CULTURAL ADAPTATION AND CALIBRATION IN SPAIN OF TWO PROMIS ITEM BANKS: PAIN-BEHAVIOUR AND PAIN-INTERFERENCE

Castejón N[1], Rodríguez A[1], Correia H[2], Choi S[2], Gálvez R[3], Mayoral V[4], Rebollo P[5]

[1]*LA-SER (BAP LA-SER Outcomes), Oviedo, Spain,* [2]*The PROMIS Statistical Center Northwestern University, Chicago, IL, USA,* [3]*Hospital Virgen de las Nieves, Granada, Spain,* [4]*Hospital de Bellvitge, Barcelona, Spain,* [5]*LA-SER (BAP LA-SER Outcomes), Oviedo, Asturias, Spain*

OBJECTIVES: To carry out the cultural adaptation and calibration in Spain of the universal Spanish version of two item banks previously developed and translated by the PROMIS Group: "Pain-Behavior-PB" (39 items) and "Pain-Interference-PI" (41 items). This will allow construction of two Computerized Adaptive Tests (CATs) for evaluating Health-Related Quality of Life (HRQoL) in patients with pain. **METHODS:** Forward and backward translations of PB and PI items were carried out. The alternative translation was then compared to the existing PROMIS translation and reviewed by 25 patients and 6 experts to evaluate the relevance and comprehensibility of items. Recommendations to revise the universal Spanish were evaluated by the PROMIS Statistical Center and by a linguist from Spain. The revised items were cognitively debriefed with 5 patients following the PROMIS *Interview Script* and discrepancies were resolved. The revised universal Spanish version of both item banks was administered to a sample of 236 subjects with chronic pain of any etiology. Unidimensionality and local independence were evaluated. The calibration of the items was done using the Samejima's graded response model. **RESULTS:** The process of cultural adaptation of both item banks for use in Spain resulted in the amendment of 8 items (5 PB -3 PI). Unidimensionality and local independence of items of both banks were confirmed. Discrimination parameters ranged from moderate to very high in almost all items of both banks. Differences according to gender were not statistically significant in PB bank ($F_{(1,225)}$=3.24;p=0.073) nor in PI bank ($F_{(1,224)}$=2.072;p=0.151). Despite the relatively small sample size, the standard errors for the item parameters are within an acceptable range (<1),with the exception of three extreme threshold parameters. **CONCLUSIONS:** The universal Spanish PROMIS Pain Behavior and Pain Interference have been calibrated in Spain and two CATs have been built to evaluate HRQoL of patients with pain in daily clinical practice.

PRM95

DOES THE CAT QUESTIONNAIRE PRODUCE SIMILAR RESULTS WHEN SELF- OR INTERVIEWER-ADMINISTERED?

Badia X[1], Agusti A[2], Soler JJ[3], Molina J[4], Cenoz S[5], Roset M[6], Garcia-Losa M[7]

[1]*IMS Health, Barcelona, Barcelona, Spain,* [2]*Hospital Clínic IDIBAPS, Barcelona & CIBER Enfermedades Respiratorias (Spain), Barcelona, Spain,* [3]*Hospital de Requena, Requena (Valencia), Spain,* [4]*Centro de Salud Francia, Madrid, Spain,* [5]*GlaxoSmithKline, Tres Cantos, Madrid, Spain,* [6]*IMS Health, Barcelona, Spain,* [7]*IMS Health, Madrid, Spain*

OBJECTIVES: The COPD Assessment Test (CAT) assesses the impact of chronic obstructive pulmonary disease (COPD) on health status in clinical practice. We examined whether the mode of administration (self vs. interviewer) influences the CAT score and/or its psychometric properties in a heterogeneous COPD population (during clinical stability and during exacerbations). **METHODS:** Observational study in 49 Spanish centers. Patients hospitalized because of exacerbations of COPD (ECOPD; n=224) and clinically stable patients (n=153) completed the CAT and other measures, including the Saint George's Respiratory Disease Questionnaire (SGRQ) and the London Chest Activities of Daily Living (LCADL) instrument. In order to replicate real-life, the CAT was self-administered (CAT-SA) or administered by health care personnel (CAT-IA) as decided by clinicians. Multiple linear regression analysis was used to determine whether mode of administration affected scores after adjusting for differences between groups. The instrument's psychometric properties were compared between groups. **RESULTS:** Of 377 patients included, 118 (31.2%) completed the questionnaire by self-administration and 259 (68.8%) by interview. Multiple regression analysis showed that the mode of administration did not affect CAT scores. Psychometric properties were good whichever mode of administration was used. Internal consistency coefficients (Cronbach's alpha: 0.86 for CAT-SA and 0.85 for CAT-IA) and test-retest reliability (intraclass correlation coefficients of 0.83 for CAT-SA and CAT-IA) were high. Correlations with SGRQ and LCADL were moderate to strong for both groups, though only the CAT-IA correlated significantly with clinical measures of COPD. Similar results were observed when testing longitudinal validity. **CONCLUSIONS:** The mode of administration does not influence CAT scores and only minimally influences its psychometric properties, suggesting that data obtained using different modes of administration can be pooled or compared. Further research is required to determine whether sensitivity to change is affected by mode of administration.

PRM96

RELIABILITY OF A PATIENT-REPORTED ADVERSE DRUG EVENT QUESTIONNAIRE

de Vries ST, de Zeeuw D, Haaijer-Ruskamp FM, Denig P

University of Groningen, University Medical Center Groningen, Groningen, The Netherlands

OBJECTIVES: Previously, a questionnaire was developed to collect information re-

garding the type, nature, and causality assessment by patients of adverse drug events (ADEs). It contains a checklist with 252 ADEs within 16 body categories. We tested the 1) impact of using this body categorization on ADE reporting, 2) test-retest reliability, and 3) feasibility of questionnaire completion. **METHODS:** Patients using glucose-lowering drugs were selected from four pharmacies. Consenting patients received the digital questionnaire twice (one week in between). Patients were randomly divided in three groups. Group 1 received the questionnaire with body categories at T0 and without categories at T1; for group 2 this was reversed. Group 3 received the questionnaire with categories twice. Agreement was calculated by Intraclass Correlation Coefficients (ICC) for reporting 1) any ADE, 2) any ADE at body category level, and 3) a specific ADE. Feasibility was based on the time needed to complete the questionnaire at T0, testing for differences using Mann-Whitney U tests. **RESULTS:** Currently, 125 patients completed the questionnaire twice. At T0, 26-27% reported an ADE using the questionnaire with categorization compared to 23% without categorization. Test-retest reliability showed moderate agreement for reporting any ADE or an ADE at body category level (ICC: 0.46 and 0.51), and poor agreement at specific ADE level (ICC: 0.38). With-without categorization showed moderate agreement for reporting any ADE (ICC: 0.63 and 0.51) but poor agreement at lower levels (ICC: 0.24-0.21 and 0.26-0.15). Overall, the median duration for questionnaire completion was 19 minutes (no difference between versions, $Z=-0.223$, $P=0.824$), and 57 minutes for those who reported ADEs (no differences, $Z=-1.402$; $P=0.161$). **CONCLUSIONS:** Use of a body categorization structure in a checklist-based questionnaire affects patients' reporting of ADEs. Without categorization, less patients tend to report ADEs. Test-retest reliability was acceptable at category level. Feasibility did not differ between the questionnaires.

PRM97

TRANSLATION AND VALIDATION OF OSTEOPOROSIS KNOWLEDGE TOOL INTO MALAYSIAN VERSION AMONG TYPE 2 DIABETICS PATIENTS

Abdulameer SA[1], Syed Sulaiman SA[1], Hassali MA[1], Subramaniam K[2], Sahib MN[1]
[1]*Universiti Sains Malaysia, Minden, Penang, Malaysia,* [2]*Penang General Hospital, George Town, Penang, Malaysia*

OBJECTIVES: To translate and examine the psychometric properties of the Malaysian version of the Osteoporosis Knowledge Tool (OKT-M) among type 2 diabetes patients and to determine the best cut-off value with optimum sensitivity and specificity. **METHODS:** The OKT English version was translated and validated using the internationally accepted and recommended methodology, which was then validated with a convenience sample of 250 T2DM outpatients. All data were collected from the Penang General Hospital, Penang, Malaysia. Instruments consisted of the Malaysian version of OKT-M and a socio-demographic questionnaire. The sensitivity and specificity of OKT-M was calculated using receiver operating characteristic curve analysis (ROC). Validity was confirmed using face (Fleiss' kappa), content (Lawshe's quantitative approach) and item analysis. Reliability was assessed using Cronbach's alpha and test–retest. **RESULTS:** The mean score of the OKT-M was 11.35± 4.21 with a mean age of 61.88±9.86 years (range: 38-90 years), and the majority were Chinese (n = 119, 47.6%). The Fleiss' kappa, content validity ratio range and content validity index values were 0.66, 0.75-1 and 0.87, respectively. Internal consistency and test–retest reliability values were 0.72 and 0.85, respectively. The mean difficulty factor and discriminatory power values were 0.47±0.16 and 0.96, respectively. The cut-off point of the OKT-M to predict osteoporosis/osteopenia was 14 with optimal sensitivity (84.1%) and specificity (85.5%). The positive and negative predictive value were 85.3% (95% CI 0.77-0.91) and 84.32% (95% CI 0.76-0.90), respectively. The area under the curve for the OKT-M was 0.92 (95% CI 0.87-0.96). By applying the cut-off point 76.4% of the T2DM patients show a low OKT-M score with mean score 9.27±3.16. **CONCLUSIONS:** The findings of this validation study indicated that the OKT-M is a reliable and valid tool with good psychometric properties in the Malaysian setting.

PRM98

OPTIMIZING DISCRETE CHOICE EXPERIMENT DESIGNS USING SIMULATIONS

Tapager IW, Jensen HH, Mønsted C, Bøgelund M
Incentive Partners, Holte, Denmark

OBJECTIVES: Discrete choice experiments (DCE) range prominently among the applied methods to elicit preferences in the field of health economics. With ongoing methodological learning, best practice remains a moving target. It is seldom plausible to implement a full factorial design and while there have been developed software tools to improve design efficiency there remains "design error". The aim of this study was to illustrate how simulation studies can inform the designing of DCE's and minimize design error given study size constraints and prior knowledge on preferences. **METHODS:** We specified a hypothetical set of attributes and levels for a DCE game as well as an expected linear additive utility function for individuals. We used Monte Carlo simulations - programmed in SAS 9.2 - to examine how different design decisions affected design error given the specified utility function, attributes and levels. **RESULTS:** Using blocking to increase choice sets minimizes design error. Maximizing the number of respondents may improve estimation but will not markedly improve design error unless used to include more choice sets. Prior knowledge – either theoretical or from prior studies – can be used to deselect choice sets that are implausible or with none or limited informational gain improving design efficiency. **CONCLUSIONS:** Simulations can provide a tool for optimizing design choices. We illustrate how it can supplement software design routines and provide an intuitive understanding of design properties and how it will likely affect the design efficiency to e.g. include more respondents, blocks or questions or use prior knowledge. Simulations are not reality – the "respondents" behave the way they are specified to behave. However, this methodology enables the researcher to isolate effects and we believe that our simulation framework can be a useful tool for practitioners to think systematically about DCE design decisions given actual study characteristics and constraints.

PRM99

TRANSLATION AND VALIDATION OF OSTEOPOROSIS SELF-EFFICACY SCALE INTO MALAYSIAN VERSION AMONG TYPE 2 DIABETICS PATIENTS

Abdulameer SA[1], Syed Sulaiman SA[1], Hassali MA[1], Subramaniam K[2], Sahib MN[1]
[1]*Universiti Sains Malaysia, Minden, Penang, Malaysia,* [2]*Penang General Hospital, George Town, Penang, Malaysia*

OBJECTIVES: To translate and examine the validity and reliability of the Malay version of the Osteoporosis Self- Efficacy Scale (OSES-M) among type 2 diabetes mellitus (T2DM) outpatients and to determine the best cut-off value with optimum sensitivity and specificity. **METHODS:** A standard "forward-backward" translation procedure was used to create the Malaysian version of the OSES-M from the original English version, which was then validated with a convenience sample of 250 T2DM outpatients. All data were collected from the Penang General Hospital, Penang, Malaysia. Instruments consisted of the Malaysian version of OSES and a socio-demographic questionnaire. The sensitivity and specificity of the OSES-M was calculated using receiver operating characteristic curve analysis. Validity was confirmed using face (Fleiss' kappa), content (Lawshe's quantitative approach) and construct validity (factor analysis). Reliability was assessed using Cronbach's alpha and corrected item-total correlations between the scales and their corresponding items. Spearman's rank correlation used to assess test–retest reliability. **RESULTS:** By employing the recommended scoring method, the mean score of OSES-M was 731.74± 197.15. Fleiss' kappa, content validity ratio range and content validity index were 0.99, 0.75-1 and 0.96, respectively. Two factors were extracted from exploratory factor analysis and were confirmed through confirmatory factor analysis. Internal consistency and test–retest reliability were 0.92 and 0.86, respectively. The optimum cut-off point of OSES-M to predict osteoporosis/osteopenia was 858 with 85% sensitivity (95% CI 0.76-0.9) and 74.5% (95% CI 0.65-0.82) specificity. The area under the curve for OSES-M in identifying osteoporotic subjects was 0.86 with 95% CI 0.8-0.92 ($P< 0.01$). **CONCLUSIONS:** The findings of this study indicate that the OSES-M is a valid and reliable instrument for measuring osteoporosis self-efficacy in the Malaysian clinical setting and research practice.

PRM100

CONVERTING EORTC QOL-C30 SCORES TO UTILITY VALUES: IS IT PLAUSIBLE?

Salek MS[1], Schmidt-Rimpler C[2], Wimberger P[3], Gilet H[4], Parsons SL[5]
[1]*Cardiff University, Cardiff, UK,* [2]*Fresenius Biotech GmbH, München, Germany,* [3]*University of Duisburg-Essen, Essen, Germany,* [4]*MAPI Consultancy, Lyon, France,* [5]*Notthingham University Hospitals NHS Trust, Nottingham, UK*

OBJECTIVES: The EQ-5D is a widely used generic preference-based measure (PBM) to derive Quality-Adjusted-Life-Years (QALYs) for use in economic evaluations. Such generic measures of health-related quality of life (HRQoL) could be insensitive for some medical conditions such as cancer. The European Organization for Research and Treatment of Cancer Quality-of-Life Questionnaire (EORTC QLQ-C30) is a widely used, non-PBM to assess the QoL in cancer patients. Although EORTC provides supplementary information for an economic evaluation, it does not produce a single QoL utility score such as that of the EQ-5D which can be used in economic analysis. Mapping is a technique to estimate the relationship between PBM (EQ-5D) and non-PBM (EORTC) to derive a single utility value. The objective of this study was to map the EORTC data from AC-01 trial onto the EQ-5D by identifying the most appropriate mapping algorithm in the literature. **METHODS:** A literature review of studies presenting an algorithm enabling utility values to be derived from EORTC data was conducted. The retrieved algorithms were compared in terms of study design, population, methodology, EORTC items/dimensions included in the final algorithm, and predictive performance. An EORTC-based utility value was calculated using the best algorithm identified. **RESULTS:** Algorithms were extracted from six sources. The algorithm reported by Rowen et al. (2011) was considered as the most appropriate when robustness of the methodology and the comprehensive nature of the dimensions compared. The EORTC-based mean utility scores calculated were 0.67 (SD= 0.13) and 0.64 (SD= 0.14) in the catumaxomab and control group at screening, respectively, and 0.72 (SD= 0.14) for the catumaxomab group at week 4. **CONCLUSIONS:** The mapping algorithm developed by Rowen et al. (2011) enabled the whole possible range of patients' health-status to be measured in the study. The derived utilities enable EORTC data to be used in economic evaluations.

PRM101

A THINK ALOUD STUDY COMPARING THE VALIDITY AND ACCEPTABILITY OF DISCRETE CHOICE AND BEST WORST SCALING METHODS

Whitty JA[1], Walker R[2], Golenko X[1], Ratcliffe J[3]
[1]*Griffith University, Meadowbrook, Queensland, Australia,* [2]*Flinders University, Adelaide, South Australia, Australia,* [3]*Flinders University, Daw Park, South Australia, Australia*

OBJECTIVES: To provide insights into the validity and acceptability of the Discrete Choice Experiment (DCE) and profile case Best Worst Scaling (BWS) methods for eliciting preferences for health care. **METHODS:** A convenience sample (N=24) undertook a traditional DCE and a BWS choice task as part of a wider survey on Health Technology Assessment decision criteria. A 'think aloud' protocol was applied, whereby participants verbalized their thinking while making choices. Internal validity and acceptability were assessed through a thematic analysis of the decision-making process emerging from the qualitative data and a repeated choice task. **RESULTS:** The think aloud data demonstrated clear evidence of 'trading' between multiple attribute/levels for the DCE, and to a lesser extent for the BWS task. For the BWS task, some participants found choosing the worst attribute/level con-