mon use across and between conditions. METHODS: Subjects were recruited by web posting and telephone screening. Those self-reporting active treatment for one of the following: Low Back 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 used a total of 53 descriptors, with 63% (n=45) of the responses from female, 60% 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, ACHELING, 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 descriptors (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 (86%), and SHOOTING (74%). OA and RA patients tended to be more similar with CONSTANT (66%/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.

PM92 EXPLORATORY MAPPING OF THE LUPUSQOL TO THE SF-6D
Meacock R1, Harrison M2, Mccolline E2, Iensenberg D2, Ferenkhh-Koroma A3, Ahmad Y3, Bouhassira D4, Sheinerman J5, Gordon C3, Griffiths B5, Maddison P8, Aki M2, Abbott J1, Teoh S2
1University of Manchester, Manchester, UK, 2Royal Blackburn Hospital, Blackburn, UK, 3University of Manchester, Manchester, UK, 4Clatterbridge Cancer Centre, Wirral, UK, 5London School of Hygiene, London, UK, 6Royal College of London, London, UK, 7Benoit Cadaval Hospital University Health Board, Llandudno, UK, 8Aberystwyth District General Hospital, Machynlleth, UK, 9University of Birmingham, Birmingham, UK, 10Freeman Hospital, Newcastle, UK, 11University of Derby, Derby, UK, 12Royal Hallamshire Hospital, Sheffield, UK, 13University 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 290 systematic lupus erythematosus (SLE) patients completed the LupusQol and SF-6D at the same assessment. Models of the relationship between the two measures 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 square 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 (R²=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 scores. Data. Potentially this could reduce patient burden if errors are lower than for many published mapping functions, signifying that the SF-6D could be used to measure health-related quality of life in SLE 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.

PM93 INFANTILE HEMANGIOMA ON FAMILIES, CREATION OF A SPECIFIC BURDEN QUESTIONNAIRE
Boccara O1, Meni C1, Ladreze C2, Bodemer C3, Vossard J4, Tarch C5
1Hôpital Necker, Paris, France, 2Hospital Pellegrin, Bordeaux, France, 3University Hospital of Toulouse, Toulouse, France, 4Benoit Cadaval University Health Board, Llandudno, UK, 5University of Manchester, Manchester, UK

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 IH. METHODS: A “Hemangioma Family Burden” questionnaire (HFB) consisting of 22 items. It was distributed accompanied by SF12 and PGBW, to obtain internal and external validation. RESULTS: The final version of the 63 questionnaire items was returned, completely or in part. The mother (63.27%) or the father (36.7%). At the time of the survey, the average age of the child with IH was 9.3±4.7 months. The survival rate was measured by Cronbach’s alpha (0.95), reflecting a good homogeneity of the 22 items. While the score of the questionnaire is expressed in two parts 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 highest extent, and 27.7±16.97 for a moderate extent. This confirms the sensitivity of the HFB. CONCLUSIONS: Internal and external validation was performed. 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.

PM94 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‐BEHAVIOR AND PAIN‐INTERFERENCE CASTIGLIONI N 1, Rodrigues A 2, Correia H 2, Choi S 2, Bellvitge, Barcelona, Spain, 3Hospital Pellegrin, Bordeaux, France, 4Hospital Virgen de las Nieves, Granada, Spain, 5Hospital de las Nieves, Granada, Spain, 6LA SERRA (BAP LA SERRA Outcomes), Oviedo, Spain, 7The PROMIS Statistical Center Northwestern University, Chicago, IL, USA, 8Hospital Virgen de las Nieves, Granada, Spain, 9Hospital de las Nieves, Granada, Spain, 10LA SERRA (BAP LA SERRA 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-FB’ (39 items) and ‘Pain-Interference-FI’ (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 FB and FI items were carried out. The alternative translation was then compared to the existing PROMIS translation and re- evaluated 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 FB – 3 FI). 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 FB bank (9.220.3.240; p=0.073) nor in FI bank (221.240; 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.

PM95 DOES THE CAT QUESTIONNAIRE PRODUCE SIMILAR RESULTS WHEN SELF- OR INTERVIEWER-ADMINISTERED?
Badía X1, Agustí A2, Soler J3, Molina J4, Cenzis S5, Rosset M6, García-Losa M7
1Hospital Clinic IDIBAPS, Barcelona, Barcelona, Spain, 2Hospital Universitari Vall d’Hebron, Barcelona, Spain, 3Centro de Salud Francia, Madrid, Spain, 4ClínicaSantKliment, Tres Cantos, Madrid, Spain, 5Hospital Materno, Barcelona, Spain, 6Hospital de las Nieves, Granada, Spain, 7Hospital Pellegrin, Bordeaux, France

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 scores and 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.

PM96 RELIABILITY OF A PATIENT-REPORTED ADVERSE DRUG EVENT QUESTIONNAIRE
de Vries SJ1, de Zeeuw D, Huijzer-Ruukamp FM, Derijg P1
1University 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. Consent- ing 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 35-37% without categorization. Test-retest reliability was 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-out 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 0.26-0.19). 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, the questionnaire response 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

Abdulmutalib S1, Syed Sulaiman S1, Hassali MA2, Subramaniam K3, Sahib MN4
1Universiti Sains Malaysia, Minden, Penang, Malaysia, 2Penang 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, George Town, Penang, Malaysia. 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. Internally consistency and test-retest reliability values were 0.72 and 0.85, respectively. The mean difficulty factor and discriminatory power values were 0.47 and 0.96, respectively. The cut-off point of the OKT-M to predict osteoporosis/osteohipia 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 0.89-0.95) with a cut-off score of 14. 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

Taggar LW1, Jensen WH, Mansted C, Reglund M
1Incentive 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 extremely desirable to implement a full factorial design and while there have been developed methodological learning, best practice remains a moving target. It is seldom plausible, whereby participants verbalized their thinking while making choices. Inter- nal 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’ be- tween 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-