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-