OBJECTIVES: Given that the number of preference-based instruments has increased over the past years, it is important to compare their performance. This work aims to study the performance of the SF-6D and EQ-5D in a variety of chronic health conditions.

RESULTS: After discarding the SF-6D data for 28 citations that were not screened, and 833 full papers were retrieved for consideration. Eight out of 145 studies were included, which focused on either diabetes or heart disease. From these, 21 utility values were identified and included in the database. Most of the excluded studies used EQ-5D or EQ-SD, but only for self-reported, non-narcotic analgesics, and did not calculate the utility index. CONCLUSIONS: Few studies were available providing utilities elicited from the Spanish population for prevalent diseases such as diabetes or heart disease. As next steps, other diseases will be explored, searches will include other drugs databases, and all experts will be contacted to identify additional, relevant information, and included studies will be quality-assessed. This ongoing review and the database of Spanish utilities will be useful for researchers developing economic evaluations or requiring information on quality of life derived from the Spanish population. The project will also allow us to identify data gaps for which further research is needed.

PRM34 EXAMINING VARIATIONS IN ITEM STRUCTURE AND CONTENT IN PRO INSTRUMENTS, OR, THERE MUST BE 50 WAYS TO EXPRESS YOUR DISTRESS

CONCLUSIONS: Detection of differences between groups in almost all conditions. Patients’ self-reporting of drug exposure is subject to memory errors and varying degrees of bias. Utilisation of prescription records is often impaired by over-prescription and use of over-the-counter (OTC) drugs. Our study compared and classified items by concept.

METHODS: The structure of each item is characterized as an item stem with a core concept, with an implicit or explicit context (e.g., a disease), event (e.g., “felt frustrated or impatient”), and stimulus (e.g., “about your symptoms”); recall period and response options. Concepts were classified using the WHO International Classification of Functioning (ICF). Similarities and differences across instruments within disease area are analyzed.

RESULTS: We decomposed over 600 items in at least 23 instruments across 5 disease areas and several generic instruments, capturing and classifying each aspect of the structure of each item. Most physical function items could be matched with specific 3-4 digit ICF codes; most emotional function items could not be matched as specifically. There was considerable variation across instruments regarding the explicit statement of context as well as the presence of a stimulus. We observed at least 6 different recall periods ranging from an implied present to “in the past year” to “in 10 years” with distinct patterns by disease area. We observed at least 9 types of response options, but the majority of items used 5-point scales. CONCLUSIONS: There are some commonalities but little standardization in how items are structured, within or across diseases. Classification and comparison of items and evaluating comparative content validity is complicated by the variation in most aspects of how the items are phrased.

PRM35 AGREEMENT BETWEEN PATIENTS’ SELF-REPORT AND PHYSICIANS’ PRESCRIPTIONS ON DRUGS AND VACCINE EXPOSURE: THE INTERNATIONAL PGRX DATABASE EXPERIENCE

OBJECTIVES: Patients’ self-reporting of drug exposure is subject to memory errors and varying degrees of bias. Utilization of prescription records is often impaired by over-prescription and use of over-the-counter (OTC) drugs. Our study compared and classified items by concept.

METHODS: The structure of each item is characterized as an item stem with a core concept, with an implicit or explicit context (e.g., a disease), event (e.g., “felt frustrated or impatient”), and stimulus (e.g., “about your symptoms”); recall period and response options. Concepts are classified using the WHO International Classification of Functioning (ICF). Similarities and differences across instruments within disease area are analyzed.

RESULTS: We decomposed over 600 items in at least 23 instruments across 5 disease areas and several generic instruments, capturing and classifying each aspect of the structure of each item. Most physical function items could be matched with specific 3-4 digit ICF codes; most emotional function items could not be matched as specifically. There was considerable variation across instruments regarding the explicit statement of context as well as the presence of a stimulus. We observed at least 6 different recall periods ranging from an implied present to “in the past year” to “in 10 years” with distinct patterns by disease area. We observed at least 9 types of response options, but the majority of items used 5-point scales. CONCLUSIONS: There are some commonalities but little standardization in how items are structured, within or across diseases. Classification and comparison of items and evaluating comparative content validity is complicated by the variation in most aspects of how the items are phrased.

PRM36 COMPARING THE PERFORMANCE OF THE SF-6D AND EQ-5D ACROSS DISEASES

OBJECTIVES: The objective of the current study was to compare the accuracy of different methods to estimate health state utility values (HSUVs) for comorbid health conditions. METHODS: Data collected during five rounds of the Welsh Health Survey were analyzed for a wide range of chronic health conditions.

RESULTS: The mean SF-6D scores for the subgroups with comorbidities ranged from 0.465 to 0.607. The mean SF-6D scores had a mean absolute error (MAE) of 0.056 and 0.121 and just 15% of estimated values were within the minimum important difference (MID) for the SF-6D (0.041) respectively. While the multiplicative method also tended to underestimate the actual SF-6D scores (MAE 0.075) it performed better when estimating MID scores in below 0.50% and 47% of estimated values were within the MID. A linear model obtained by mapping the disutilities associated with the mean SF-6D scores for two subgroups with single conditions (plus the interaction between the two) onto the mean SF-6D scores for subgroups with comorbidities gave the most accurate results overall. The predicted SF-6D scores had a mean absolute error of 0.0191 and 88% of predicted SF-6D scores accurate to within the MID.

CONCLUSIONS: While in our data the linear model gave the most accurate results, additional research is required to validate our results.