OBJECTIVES: The quality-adjusted time without symptoms and toxicity (Q-TWiST) methodology has been used to assess the clinical benefits (prolonged [progression-free] survival) and costs (toxicities) of oncology therapies. This study was conducted to systematically review and quantitatively summarize published Q-TWiST assessments of cancer treatments. **METHODS:** A systematic search and review was conducted in MEDLINE to identify original studies reporting the Q-TWiST information—including time with toxicity (TOX), time before disease progression without toxicity (TWiST), and time in relapse after disease progression (REL)—for all oncology treatment groups, as available. Utilities for Q-TWiST were also captured; when a base case for utilities was not selected in a study, the following was assumed: u(TWiST)=1, u(REL)=0.5, and u(TOX)=0.5. The relative gain in Q-TWiST for active treatment arms was calculated as the difference in Q-TWiST divided by mean overall survival of control arm. Relative gains ≥10% and ≥15% were considered to be a clinically important and clearly clinically important difference, respectively. **RESULTS:** Upon review of 84 initially identified articles, 39 were excluded for the following reasons: no Q-TWiST was reported (n=22), not oncology-related (n=13), other reasons (n=4). Forty-five studies were included and reported a total of 69 Q-TWiST comparisons across 10 cancers. The most commonly used utilities for Q-TWiST calculation were u(TWiST)=1, u(REL)=0.5, u(TOX)=0.5 (n=28, 62.2% of articles). Using base-case utility values, the mean (range) Q-TWiST gain was 5.2 (-6.8 to 61) months and the mean relative gain was 9.0% (-13.3% to 60.0%); 41.8% and 17.9% of studies reported relative gains ≥10% and ≥15%, respectively. Applying u(REL)= u(TOX)=0.5 to comparisons with sufficient data (n=65), the mean standardized Q-TWiST gain was 5.4 (-4.1 to 61) months and mean standardized relative gain was 8.2% (range -15.0% to 54.5%). **CONCLUSIONS:** This review of Q-TWiST for cancer therapies can serve as benchmark against which future analyses can be compared.

PRM80

EQ-5D HEALTH UTILITIES ARE ESTIMATED SUBJECT TO CONSIDERABLE UNCERTAINTY

Pullenayegum EM[1], Xie F[2]
[1]*Hospital for Sick Children, Toronto, ON, Canada,* [2]*McMaster University, Hamilton, ON, Canada*

OBJECTIVES: The EQ-5D is often used to measure health utilities. Uncertainty in the EQ-5D scoring algorithm is routinely ignored. We aim to quantify the extent of uncertainty in the US EQ-5D-3L scoring algorithm, which was based on data from 3773 respondents -- largest valuation study to date. **METHODS:** We re-fitted the US scoring algorithm using the same data and functional form as was originally used, omitting each health state in turn and examining the error in the predicted mean utilities. We then used a mixed effects model, including a random effect for health state, adopting a Bayesian perspective to estimate the predictive distribution of the mean utilities for health states not included in the valuation study (which captured 43 of 243 health states). This allowed us to estimate uncertainty in the scoring algorithm. **RESULTS:** The mean absolute error for predicted mean utilities on cross-validation was 0.033; the mean absolute error for a perfect model, accounting for sampling error in the observed mean utilities, would have been 0.01. The root mean squared error was 0.042; for a perfect model it would have been 0.013. The standard deviation for the random effect for health state was 0.03, suggesting that the width of the confidence interval for the mean utility for a randomly selected health state is around 0.12. The Bayesian model indicated that the width of the 95% credible interval for the mean utilities varied from 0.015 to 0.45, with a median width of 0.18 and interquartile range of 0.15 to 0.22. **CONCLUSIONS:** EQ-5D health utilities are subject to considerable uncertainty (for comparison, the MID for EQ-5D utilities is 0.05 to 0.08). Other countries' scoring algorithms are based on smaller sample sizes and so subject to greater uncertainty. This uncertainty should be accounted for when using EQ-5D health utilities in economic evaluations.

PRM81

LAM EMPLOYMENT ABSENCE AND PRODUCTIVITY SCALE (LEAPS): FURTHER VALIDATION STUDIES IN MAJOR DEPRESSIVE DISORDER

Lam RW
University of British Columbia, Vancouver, BC, Canada

OBJECTIVES: There is increasing clinical and research interest in monitoring functional outcomes, including occupational functioning, in people with Major Depressive Disorder (MDD). Self-rated scales are most practical for clinical settings, but there are few work functioning scales that have been validated in depressed populations. We now report further psychometric studies of the Lam Employment Absence and Productivity Scale (LEAPS), a brief 10-item self-report questionnaire. **METHODS:** Patients meeting DSM-IV criteria for MDD completed the LEAPS during initial assessment and following various treatment protocols in specialist outpatient and family practice settings, and other scales including the self-rated Sheehan Disability Scale (SDS) and Health and Work Performance Questionnaire (HPQ), and the clinician-rated Clinical Global Impression-Improvement (CGI-I) scale. Standard statistical analyses for scale validation were conducted. **RESULTS:** In a sample of 418 patients at baseline assessment, internal consistency was high, with Cronbach's alpha of 0.93 (for the LEAPS total score) and 0.89 (for the productivity subscale). Convergent validity was supported by significant correlations of the LEAPS total score and productivity subscale with the SDS work/role item score (r=0.49, 0.43, respectively, both p<0.0001) and the HPQ Overall Performance score (r=-0.42, -0.53, respectively, both p<0.0001). In a sample of 104 patients repeating the LEAPS within a 1 week period, excellent test-retest correlations were found (r=0.95, p<0.0001). In a sample of 176 patients assessed before and after 8 weeks of antidepressant treatment, the LEAPS showed good responsivity to change, with differences in means between CGI-I categories of Very Much Improved (LEAPS total score change =10.7±5.5; productivity subscale change=3.9±2.5), Much Improved (total=8.3±4.5; subscale=3.0±2.3) and Minimally Improved/No Change/Worse (total=4.5±5.7; subscale=1.4±2.9); differences were significant for both LEAPS total score change (ANOVA, F=7.29, df=2,83, p=0.001) and for productivity subscale change (F=4.07, df=2,83, p=0.021). **CONCLUSIONS:** These data support the LEAPS as a valid, reliable and responsive instrument in patients with MDD.

PRM82

USING VISUAL AIDES IN TRANSLATION OF A SPASTICITY SYMPTOM QUESTIONNAIRE

Brandt BA[1], Talbert M[2], McKown S[3], Gawlicki MC[1], Bernardi K[3]
[1]*Corporate Translations, Inc., East Hartford, CT, USA,* [2]*Corporate Translations Inc., Chicago, IL, USA,* [3]*Corporate Translations, Inc., Chicago, IL, USA*

OBJECTIVES: The objective of this study was to determine the effectiveness of visual aids for translation of the Spasticity Symptom Assessment-Upper Limb (SSA-UL) questionnaire as an additional support tool for linguists. Some items in the SSA-UL concern specific physical positions experienced by stroke patients, due to muscle spasticity. It was theorized that translation would be improved by providing translators with a visual of physical positions, supplementing a textual definition, which may be complex and open to varied interpretations. **METHODS:** The SSA-UL was translated into 14 languages. Translators were provided a textual definition and images of the following physical positions: (1) *hand clenching*, (2) *finger clenching*, (3)*hand curling* and (4) *finger curling*. Back-translations were analyzed for conceptual equivalency with the source text. In addition, each translator (n=28) was given a questionnaire about the effectiveness of the visual aids. **RESULTS:** Translations of *hand clenching* and *finger clenching* were conceptually equivalent to the source 100% of the time. For both *hand* and *finger curling*, only French-Canada and Polish-Poland struggled to find an equivalent translation, ultimately achieving a conceptually equivalent translation. German-Germany linguists had initially translated *curling* and *clenching* incorrectly, but utilized the images to revise the translation. Nineteen (19) linguists responded to the questionnaire. Fifteen (79%) found the visual aids helpful. For comparison, previous translations of other physical positions translated without visual aids were analyzed with a similar language sample. *Bending* was translated as conceptually equivalent to the source 83% (10/12) of the time, *stooping* 58% (7/12) and *feet dropping forwards* 50% (6/12). **CONCLUSIONS:** Back-translations showed that visual aids added value to the translation process. Linguists surveyed reported that the visual aids assisted them in finding the correct terminology for the physical positions. Moving forward, translation of questionnaires containing items about physical positions may benefit from visual aids.

PRM83

PSYCHOMETRIC PROPERTIES OF THE BRIEF FATIGUE INVENTORY-SHORT FORM IN SYSTEMIC LUPUS ERYTHEMATOSUS

Al Sawah S[1], Tomaszewski EL[2], Gemmen E[3], Naegeli A[1]
[1]*Eli Lilly and Company, Inc, Indianapolis, IN, USA,* [2]*Quintiles, Pittsburgh, PA, USA,* [3]*Quintiles, Rockville, MD, USA*

OBJECTIVES: A 12-week prospective, observational study was conducted to evaluate the psychometric properties of the Brief Fatigue Inventory-Short Form (BFI-SF) and determine its appropriate use and performance for the measurement of fatigue in systemic lupus erythematosus (SLE) clinical trials. **METHODS:** Participants ≥18 years, who self-reported a physician diagnosis of SLE (confirmed by medical record review) and active SLE demonstrated by a Systemic Lupus Activity Questionnaire (SLAQ) score ≥11 (0-44 scale), were recruited using a free electronic medication monitoring service. All participants completed the BFI-SF, Multidimensional Assessment of Fatigue (MAF), and Short Form-36 (SF-36) electronically at baseline, week 2, and week 12. Score distributions, internal consistency, test-retest reliability, and construct validity were evaluated. **RESULTS:** A total of 122 participants were included in the study. The mean age was 45.7 years, 95.9% were female, and 68.9% were non-Hispanic white. Cronbach's alpha were >0.9 for all BFI-SF items. Test-retest reliability of the BFI-SF showed a stable intraclass correlation for item #7 (ICC 0.76), and BFI domain scores had higher correlations (around 0.5) than most items (around 0.3-0.4). Construct validity was measured by strength of the correlations of the BFI-SF severity domains and global scores, and was moderately positively correlated to the SLAQ score (r>0.4). The domain and global scores were moderately negatively correlated to the SF-36 Vitality and Physical Function domains and SF-36 Physical Component score (r<-0.3). The BFI-SF item #3 for worst fatigue was highly positively correlated to the MAF (r=0.6). Patients with less severe fatigue (MAF≤36) scored lower than patients with more severe fatigue (MAF>36) on all domains of the BFI-SF (total score, 4.66±1.55 vs. 6.80±1.13; p<0.0001). **CONCLUSIONS:** Assessment of fatigue severity as measured by the BFI-SF demonstrated validity and reliability in a sample of patients with moderate-to-severe SLE and may be used as a patient-reported outcome tool in clinical trials.

PRM84

DEVELOPMENT AND VALIDATION OF THE FATIGUE SYMPTOMS AND IMPACTS QUESTIONNAIRE – RELAPSING MULTIPLE SCLEROSIS (FSIQ-RMS™)

Stokes J[1], Schüler R[2], Yaworsky A[1], Hunsche E[2]
[1]*Adelphi Values, Boston, MA, USA,* [2]*Actelion Pharmaceuticals Ltd, Allschwil, Switzerland*

OBJECTIVES: Patient-reported outcome (PRO) instruments commonly used to measure fatigue in multiple sclerosis (MS) have not been developed according to the 2009 FDA PRO guidance. A qualitative research study was conducted to develop a new PRO, according to the guidance, measuring fatigue symptoms and impacts in relapsing MS (RMS). **METHODS:** Adult patients with relapsing-remitting MS (RRMS) were interviewed to elicit fatigue-related symptoms and impacts. Based on spontaneously reported symptoms, a draft PRO was developed. This was debriefed in cognitive interviews with further RRMS patients, and subsequently revised. Applicability of the PRO to other RMS populations was determined in content confirmation interviews with progressive-relapsing MS (PRMS) and relapsing secondary-progressive MS (RSPMS) patients. Institutional review board approval and participant informed consent were obtained before interviews. **RESULTS:** Participants were representative of RMS patients in clinical studies and practice; most had mild-to-moderate disease. Concept elicitation and cognitive debriefing included 17 (mean±SD age 43.9±13.3 [years]; 77% female) and 20 (47.0±12.0; 80% female) RRMS patients, respectively; content confirmation included 5 PRMS (52.6±12.5; 80% female) and 5 RSPMS (52.4±10.8; 60% female) patients. Saturation of concepts was reached during concept elicitation. Cognitive debriefing confirmed that participants understood PRO instructions, items,