OBJECTIVES: There already exists an algorithm to generate utility scores from SF-36 data. This has a serious floor affect and does not accurately describe EQ-5D data. The purpose of this study was to develop an alternative model that better accommodates utility scores from the entire range of ill-health. METHODS: Data on 1946 inpatients and outpatients were abstracted from the Health Outcomes Data Repository (HODAR) in Cardiff, UK and utilised in the model building. Validation of the model was made using independent data from 554 respondents from a survey of people with diabetes in Cardiff that included the EQ-5D, SF-36. Both surveys also included complete inpatient records, diagnoses, procedures, in and out of hospital mortality, biochemistry, unit costs, inpatient medications, risk factors and demography. Various multivariate parametric and non-parametric techniques were applied to calculate an optimised model on the HODAR data set. RESULTS: When applied to the diabetes data, the calculated and actual values were as follows. Data are the minimum and maximum utility values, the inter-quartile range, median, mean, and standard deviation, respectively. Results from the survey were as follows: -0.48, 1, 0.52, 0.85, 0.71, 0.65 and 0.32, respectively. For the existing model, 0.26, 1, 0.61, 0.85, 0.74, 0.72 and 0.16, respectively. For the revised model, 0.05, 1, 0.52, 0.94, 0.76, 0.71 and 0.23, respectively. CONCLUSION: This new algorithm represents a notable improvement on the existing model. Proper evaluation of new medicines is vital, and the floor effect tends to reduce incremental cost utility ratios when achieving certain thresholds is important for drug reimbursement. HODAR will evolve to be the largest series of outcome data anywhere.

ASSESSING FUNCTIONAL STATUS IN CHILDREN: A REVIEW OF THE CHILDHOOD HAQ FOR NON-ARTHRITIS INDICATIONS
 Tran KT, Stephens JM, Gold KF, Kimura A, Pashos CL
 1Abt Associates Clinical Trials, Bethesda, MD, USA; 2Transkaryotic Therapies Inc, Cambridge, MA, USA; 3Abt Associates Clinical Trials, Cambridge, MA, USA

OBJECTIVES: To examine systematically the Childhood Health Assessment Questionnaire (CHAQ) for use in non-arthritic populations and to assess its sensitivity to effects observed in clinical trials of drug treatments. METHODS: A primary computerized literature search (1966-present) identified English-language publications with the term, “CHAQ”. Supplemental searches were conducted to identify use of CHAQ in clinical studies of specific juvenile arthropathies for comparisons. Manual searches were also conducted from the bibliographies of the articles retrieved. Articles were selected for detailed review if they met specific criteria related to the CHAQ's development and use. Articles were examined with respect to study type, country, patient population, psychometric validity in a specified patient population, and correlations with objective clinical measures, such as joint range of motion. RESULTS: Of 237 abstracts reviewed, 59 articles met the selection criteria for detailed review. Roughly three-quarters of the CHAQ articles were cross-cultural adaptation and validation studies for countries outside the U.S. The CHAQ has been utilized to measure disability and functional status in patients with juvenile rheumatoid arthritis, juvenile spondylarthropathy, spina bifida, juvenile idiopathic inflammatory myopathies, and juvenile onset mixed connective tissue disease. The CHAQ correlated with several joint range of motion measurements and scales (e.g. EPM-ROM, JAM, and GROMS), although there is some concern regarding the sensitivity of these other instruments. Future observational studies have used the CHAQ, and only one study using the CHAQ was a randomized clinical trial. CONCLUSIONS: The CHAQ appears valuable in measuring disability and functional status in populations for which it was developed and validated (i.e. juvenile arthritis), although additional research is needed to determine the instrument's ability in assessing sensitivity to change in drug therapies before use in non-arthritic populations. Future studies should be undertaken to provide empirical evidence of its reliability in populations for which the CHAQ has not been validated.

COMPARING SHORT FORM 6D, STANDARD GAMBLE, AND HEALTH UTILITIES INDEX MARK 2 AND MARK 3 UTILITY SCORES: RESULTS FROM TOTAL HIP ARTHROPLASTY PATIENTS
 Feeny DH, Wu L, Eng K
 1Institute of Health Economics, Edmonton, AB, Canada; 2University of Alberta, Edmonton, AB, Canada

OBJECTIVES: The Short-Form 36 (SF-36) is a widely used measure providing scores for 8 domains of health status. Recently Brazier et al. created the SF6D system that covers 6 of the 8 domains and provides single-summary utility scores. The first objective is to compare 6D, standard gamble (SG), Health Utilities Index Mark 2 (HUI2), and HUI Mark 3 (HUI3) scores. The second objective is to compare the change between pre- and post-surgery. The third objective is to compare the responsiveness of the 4 utility measures. METHODS: A cohort of patients referred for evaluation for total hip arthroplasty (THA) were evaluated at the time of referral and followed until 3 months after THA. Patients were assessed with the SF-36, HUI2, HUI3, the SG, and specific measures. Agreement is assessed using the intra-class correlation (ICC). Responsiveness is assessed using effect size, standardized response mean, and paired sample t-test. RESULTS: Complete data was available for 86 patients at baseline and for 63 at both pre- and post-surgery. At baseline mean 6D (0.61), SG (0.62), and HUI2 (0.62) scores were similar; HUI3 (0.52) was lower. There was less variability in 6D scores; standard deviation of 0.10 compared to 0.32 for SG, 0.19 for HUI2, and 0.22
for HUI3. At baseline, agreement between 6D and SG scores was quite low; agreement between 6D and HUI2 was moderate; and agreement between 6D and HUI3 was weak. Agreement at pre- and post-surgery was similar. The change in scores between post- and pre-surgery was lowest for 6D (0.10), intermediate for SG (0.16), and larger for HUI2 (0.22) and HUI3 (0.23). In general, responsiveness was highest for HUI3 followed by 6D, HUI2, and the SG. CONCLUSIONS: Agreement between SG scores and 6D and HUI scores was low. The estimate of change in utility associated with THA was lowest for 6D; 6D was less responsive than HUI3. Further studies that provide comparisons appear to be warranted.

**PMD33**

**RELATIONSHIP OF QUALITY OF LIFE DOMAINS TO PREFERENCE-BASED MEASURES OF QUALITY OF LIFE**

**Miller LA1, Singer ME2**

1West Virginia University School of Pharmacy, Morgantown, WV, USA; 2Case Western Reserve University School of Medicine, Cleveland, OH, USA

**OBJECTIVES:** It is generally accepted that both the physical and the mental domains contribute to quality of life (QOL). However, the extent to which the domains/dimensions contribute to overall QOL or health state utilities remains controversial. The purpose of this study was to investigate how domains of health related quality of life (HRQOL) are associated with preference-based measures of QOL. **METHODS:** We employed a cross-sectional, supervised self-administered survey to 441 Caucasian-Americans and 341 African-Americans. For one of three hypothetical health state scenarios, each person rated the condition on 4 domains of QOL: Physical Pain/Discomfort, Independence, Social Relationships, Psychological Aspects. Respondents also assessed the condition using the Visual Analog Scale (VAS, 0–100) and time trade-off (TTO). Multiple linear regression models were constructed and analyzed. The outcome variables were VAS and TTO scores. The independent variables were the 4 domain scores, adjusting for current health. **RESULTS:** The independence and physical pain/discomfort domains were strongly associated with VAS (β estimates = −8.643 and −7.138; both p < 0.0001). The Social domain had less impact (β estimate = −1.819, p = 0.0142), and the Psychological domain was not a significant predictor of VAS. In contrast, the only significant predictor of utility score (TTO) was independence (β estimate = −0.0758, p < 0.0001). **CONCLUSIONS:** The psychosocial domains are more strongly related to HRQOL as measured by VAS, than to utility scores as measured by TTO. Only the independence domain translates from HRQOL to utilities. This supports previous work that suggests that HRQOL and utilities may measure different concepts. This raises questions about the appropriateness of quality adjustment using TTO, which deals with morbidity vs. mortality, in situations where there is no mortality factor.

**PMD34**

**COMPARISON OF SF-36 SUMMARY AND PREFERENCE-BASED UTILITY SCORES ACROSS GROUPS DIFFERING IN DISEASE SEVERITY: RESULTS FROM THE MEDICARE HEALTH OUTCOMES SURVEY**

Ware JE, Bierman AS, Gandek B, Sinclair SJ, Lawrence WF

1QualityMetric Incorporated, Lincoln, RI, USA; 2Agency for Healthcare Research and Quality, Rockville, MD, USA; 3Health Assessment Lab, Boston, MA, USA

**OBJECTIVES:** To compare SF-36 psychometrically-based and preference-based utility scores across groups of elderly patients differing in disease severity. **METHODS:** Two cohorts of the Medicare Health Outcomes Survey (HOS), randomly sampled from each of 269 (1998) and 283 (1999) managed care plans, were analyzed at baseline (response rates = 60% and 67%) and 2 years (response rates = 82% and 83%). Severity of illness classifications were based on self-reported symptom severity for congestive heart failure (n = 11,192), ischemic heart disease (n = 29,666), respiratory disease (n = 19,414), diabetes (n = 25,131), and arthritis (n = 75,938). For each condition, respondents were also classified as with or without current depressive symptoms. Baseline and change scores were compared across condition severity and depressive groups using SF-36 Physical (PCS) and Mental (MCS) Component Summary scores and the SF-6D health utility index. The relative validity (RV) of the SF-36 measures was evaluated by dividing the summary statistical result (F-ratio) for each measure by the result for the best measure (RV = 1.00), for each comparative test. **RESULTS:** When classified by severity of condition (worst to least), mean scores differed substantially from 30.2–45.8 (1.6 SD units) for PCS, 40.6–51.7 (1.1 SD units) for MCS and 0.54–0.75 (1.4 SD units) for SF-6D. For groups differing only in severity of condition, PCS (RV = 0.95–1.00) and SF-6D (RV = 0.89–1.00) showed similar levels of RV, while MCS was substantially lower (RV = 0.31–0.46). For groups differing in both severity of condition plus depressive symptoms, MCS and SF-6D had the highest RV. Similar patterns of similarities and differences in RV estimates were observed across measures in longitudinal comparisons of groups that changed in condition severity and/or depressive symptoms. **CONCLUSIONS:** Psychometrically-based SF-36 summary measures may discriminate/respond best when differences are concentrated in the health component that they best measure; a utility index may discriminate/respond best when multiple components differ across severity groups or change over time.

**PMD35**

**QUALITY OF LIFE DIFFERENCES IN OLDER ADULTS WITH VARIED COMORBID CONDITIONS**

Martin ML, Bushnell DM, Budhiarso I

Health Research Associates, Inc (HRA), Seattle, WA, USA