

Abstracts

A13

compare guideline-adherent and routine clinical practice. Value of information analysis will be employed to identify areas for future research. We are applying this translational approach to various current and developing examples of PM in cancer: 1) trastuzumab for human-epidermal growth factor receptor-2 positive BC; 2) gene-expression profiling to identify patients who will benefit most from adjuvant treatment in BC; 3) cytochrome P450 2D6 testing to select patients for adjuvant tamoxifen therapy in BC; and 4) testing for Lynch Syndrome in CRC patients and their family members to inform treatment and preventative interventions. This research will develop evidence-based information for patients, providers, industry, researchers and policymakers to objectively assess how PM can be beneficial and efficient in improving cancer outcomes.

PODIUM SESSION III: RESEARCH ON THE USE OF UTILITY MEASUREMENT

ASSOCIATION BETWEEN UTILITY AND TREATMENT AMONG PROSTATE CANCER PATIENTS

Jayadevappa R, Chhatre S, Malkowicz B

University of Pennsylvania, Philadelphia, PA, USA

OBJECTIVES: To analyze the association between utility, treatment and generic and prostate-specific health related quality of life (HRQoL) among prostate cancer patients. **METHODS:** In this longitudinal cohort study we recruited 201 (≥ 45 yrs) newly diagnosed prostate cancer patients from urology clinics of an urban academic hospital. Participants completed Quality of Wellbeing (QWB-SA), generic (SF-36) and prostate-specific (UCLA-PCI) HRQoL surveys prior to treatment and up to 24 months post-treatment. Clinical and demographic data were obtained via medical chart review and utility scores were computed using QWB-SA. To analyze the relationship between treatment and utility we used linear mixed effects models, after adjusting for covariates. Similar models were used to examine association between generic and prostate-specific HRQoL and utility. **RESULTS:** Mean baseline utility was comparable between radical prostatectomy (RP) and external beam radiation therapy (EBRT) groups (0.73 vs. 0.69, $p = 0.1750$). Mixed effects models indicated that RP was associated with higher utility at 24 month (OR = 1.12, $p = 0.027$), after controlling for covariates. RP was associated with improved functioning for role physical, role emotional, vitality, mental health and bodily pain and impaired urinary function. Higher scores on generic health subscales were indicative of higher utility. Also, for prostate-specific HRQoL, higher scores on bowel function, sexual function, urinary bother and bowel bother were associated higher utility. **CONCLUSIONS:** Treatment appears to have significant association with post-treatment utility. Thus, utility assessment provides an important quantitative tool to support patient and physician clinical treatment decision making process of prostate cancer care.

UTI

COMPARISON OF HEALTH STATE UTILITY ESTIMATES IN COST-EFFECTIVENESS ANALYSIS

Šlejko JE¹, Ghushchyan VH¹, Sullivan PW²

¹University of Colorado Denver, Aurora, CO, USA, ²Regis University School of Pharmacy, Denver, CO, USA

OBJECTIVES: This study examined the effect of different utility instruments and tariffs on chronic condition utility scores and incremental cost-effectiveness ratios (ICERs). **METHODS:** Postponement of diabetes was modeled in a high-risk population receiving a hypothetical intervention, as compared to a similar control group. Utility estimates for the general population, diabetes, hypertension, stroke, myocardial infarction (MI) and congestive heart failure (CHF) were taken from catalogues developed by Sullivan et al. for United States and United Kingdom EuroQOL-5D tariffs (EQ-5DUS and EQ-5DUK) and SF-6D, using the Medical Expenditure Panel Survey. Quality-adjusted life-years (QALYs) were modeled for 20 years using a 3% discount rate. The discounted cost of the intervention was assumed to be \$150,000. **RESULTS:** The initial health state utility was lowest using the SF-6D(.709), and highest using EQ-5DUS(.800). The highest estimates for diabetes(.708), hypertension(.761) and MI(.575) were found using EQ-5DUS, but the highest stroke(.534) and CHF(.503) estimates were found with the SF-6D. The lowest estimates for stroke(.189), MI(.516) and CHF(.189) were from the EQ-5DUK, but the lowest estimates for hypertension(.583) and diabetes(.618) were from the SF-6D. The EQ-5DUK resulted in the largest marginal utility decrement for all conditions. The smallest decrement for hypertension(-.002) and diabetes(-.022) resulted from the SF-6D, but from the EQ-5DUS for stroke(-.069) and CHF(-.055). When these estimates were applied to the model, the resulting QALYs gained from the intervention were greatest from the EQ-5DUS(11.32) and least from the EQ-5DUK(9.77). Incremental QALYs gained were greatest for EQ-5DUK(3.278) and least for the SF-6D(2.880). The ICER varied from \$46,000–\$52,000 depending on the instrument/tariff. **CONCLUSIONS:** Different utility instruments/tariffs resulted in unsystematic differences in chronic condition utilities, but the marginal decrement in each of these conditions was systematically lower for the EQ-5DUK and may be a trend across all conditions. Incremental QALYs gained differed by instrument/tariff and the ICER varied from \$46,000–\$52,000.

UT2

RELIABILITY OF HEALTH UTILITIES INDEX (HUI) SCORES: PATIENT AND PARENT INTER-RATER AGREEMENT ACROSS TWO CLINICAL TRIALS OF TREATMENT FOR ACUTE LYMPHOBLASTIC LEUKEMIA (ALL) IN CHILDHOOD

Horsman JB¹, Rae CS², Furlong W¹, Silverman LB³, Sallan SE², Athale U², Barr RD²

¹Health Utilities Inc., Dundas, ON, Canada, ²McMaster University, Hamilton, ON, Canada,

³Harvard Medical School, Boston, MA, USA

OBJECTIVES: To assess differences in the reliability of HUI Mark 3 (HUI3) health-related quality of life (HRQL) utility scores for patients between self and parent assessments across two Dana-Farber Cancer Institute (DFCI) clinical trials for treatment of ALL during childhood. **METHODS:** Patients were enrolled in either the DFCI 95-001 or 00-001 trial, and were ≥ 12 years of age at the time of HUI survey. Patients and parents, blind to each other, completed HUI questionnaires at each of 5 trial phases: induction; CNS prophylaxis; intensification; continuation; and post-treatment. Reliability was assessed in terms of inter-rater agreement of individual scores and differences in mean scores. Agreement was quantified using the single-measure two-way mixed model intra-class correlation coefficient (ICC). An ICC of 0.41–0.60 represents moderate reliability, 0.61–0.80 good reliability, and 0.81–1.00 very good reliability. Mean differences of >0.03 are clinically important. Statistical significance was set at $p < 0.05$. **RESULTS:** The number of patient and parent paired assessments varied by assessment phase for both the 95-001 (minimum = 29, maximum = 50) and the 00-001 (minimum = 28, maximum = 54) trials. ICCs in the two trials ranged from 0.49 ($p < 0.05$) to 0.88 ($p < 0.05$). There was substantial overlap of ICC 95% confidence bounds across the two trials at each of the five assessment phases. There was no significant difference ($p > 0.06$) between patient-parent pairs of scores at any assessment phase in either trial. The difference between trials in mean patient-parent scores was ≤ 0.03 and insignificant ($p > 0.08$) for each of the 5 assessment phases. **CONCLUSIONS:** Agreement between patient and parent scores was moderate or better for all assessment phases in both trials. There were no important differences in mean patient and parent scores for any of the assessment phases of the two trials. Inter-rater reliability of scores was similar across the two trials. Parental assessments provide acceptable and consistent estimates of HRQL for children.

UT3

EVALUATING WILLINGNESS TO PAY THRESHOLDS FOR A DEMENTIA CAREGIVING INTERVENTION

Jukowitz E, Gitlin L, Pizzi LT

Thomas Jefferson University, Philadelphia, PA, USA

OBJECTIVES:—Assess the methodological challenges of evaluating appropriate willingness to pay thresholds (WTP) in cost-effectiveness studies which do not use a societal perspective or a QALY outcome measure. **METHODS:**—Tailored Activity Program (TAP), an intervention designed to reduce caregiver burden for dementia patients, served as a case study. Caregiver interventions such as TAP employ an individual perspective and non-QALY outcome measures such that standard societal WTP thresholds are not applicable. Two outcome measures related to caregiver burden were used 1) reduction in hours “on duty,” and 2) reduction in hours “doing things.” To estimate appropriate WTP values for each TAP outcome measure, we identified three studies which met these inclusion criteria: 1) published studies in the past 5 years using contingent valuation methodology to identify WTP, 2) assessed WTP for a dementia-related intervention that required an out-of-pocket expenditure, and 3) asked caregivers what they would be willing to pay for an outcome of reducing caregiver burden. We also assessed WTP based on the potential financial savings caregivers could achieve from purchasing TAP. To assess proportion of time TAP was cost-effective, we built a Monte Carlo simulation to test the four WTP values identified. **RESULTS:**—For the outcome measure “on duty” WTP varied between \$1.06/hr–\$4.58/hr. WTP for the outcome measure “doing things” varied between \$2.21/hr–\$9.57/hr. Applying WTP values to TAP indicates TAP cost-effectiveness varies between 50%–80% for both outcome measures. **CONCLUSIONS:**—If WTP data can not be collected prospectively or societal values can not be applied, evaluating WTP using comparable studies appears to be an acceptable method for informing decisions makers of potential cost-effectiveness. Application of WTP to TAP shows potential cost-effectiveness that can be expected under different WTP scenarios.

UT4

POSTER SESSION I

CONCEPTUAL PAPERS & RESEARCH ON METHODS – Clinical Outcomes Methods

PMCI

EVALUATING CLINICIAN REPORTED OUTCOME (CRO) ENDPOINTS FOR FDA REGULATORY APPROVALS

Nixon A, Gallop K

Oxford Outcomes Ltd, Oxford, Oxon, UK

OBJECTIVES: Clinician reported outcomes (CROs) are the most commonly observed endpoint in FDA approved product labels (Wilke et al, 2004) but few have been adequately scrutinized in terms of their suitability as endpoints. This study evaluates four widely used CROs in order to assess their suitability as endpoints for regulatory approvals. **METHODS:** Published evidence on the Karnofsky Performance Status