variability in CERs for some drugs for different indications, in some cases also varying by biomarkers. Primary care drugs had lower and less variable CERs than specialty drugs. Variations also exist in methodology used by different groups in modeling cost effectiveness, especially for time horizon and comparator. Majority of primary care drugs were modeled for a time horizon of 35-40 years or lifetime to demonstrate cost effectiveness. Among the top 10 drugs, quetiapine and erythropoietin had the highest variability across different studies, and atorvastatin, salmeterol/fluticasone and clopidogrel had the most consistent ICER values across studies. CONCLUSIONS: This analysis shows the range, variability and methods used for calculation of ICER values for these high budget impact drugs and provides lessons for executives and policy makers.

CONCEPTUAL PAPERS & RESEARCH ON METHODS – Patient-Reported Outcomes Studies

A COMPARISON OF THE DISCRIMINATIVE AND EVALUATIVE PROPERTIES OF THE SF-36 AND THE SF-6D INDEX

Muthuk A1, Bresser J, Walters S2
1University of Arizona, Tucson, AZ, USA, 2University of Sheffield, Sheffield, UK

OBJECTIVES: To examine whether the move from the SF-36 to the SF-6D entails a loss in discriminative and evaluative strength, the magnitude of that loss, and how it matters. METHODS: The study used relative validity (RV); a ratio of two F statistics, and standardized response means (SRM) to evaluate sensitivity and responsiveness of the SF-36 scales and SF-6D index. An RV of 1 reflected the most sensitive/responsive scale and the smaller the RV the less sensitive/ responsive the measure would be. Criteria was set for interpreting effect sizes which corresponded to the SRMs. The data were used initially collected for prior studies in seven diseases/conditions: chronic obstructive pulmonary disease, leg ulcers, the elderly in exercise, osteoarthritis, irritable bowel syndrome, migraine and obesity. Identified discriminative and evaluative variables were used to compare RVs and SRMs of the SF-36 scales and the SF-6D index. The mean RV differences and mean SRMs differences between the SF-36 scales and the SF-6D index represented the loss or gain in sensitivity. RESULTS: Data was available from a total of 10,089 subjects. No single SF-36 scale consistently had the largest RV or SRM, and there was no largest RV or SRM observed for the SF-6D index. The mean RV differences and mean SRMs differences between the SF-36 scales and the SF-6D index in any condition studied. Compared showed the SF-6D index was more discriminative with a mean RV difference of 0.09, (95% CI: 0.07 to 0.12) and more responsive with a mean SRM difference of 0.08, (95% CI: 0.07 to 0.16) than the SF-36 scales. However, based on longitudinal RVs the index was less responsive with a mean RV difference of 0.07, (95% CI: 0.01 to 0.13) than the SF-36 scales. CONCLUSIONS: Moving from the SF-36 to the SF-6D index entails a loss in evaluative strength and a gain in discriminative strength, a loss/gain too small to matter given the merits of this index in any condition studied. Comparisons showed the SF-6D index was more discriminative and evaluative with a mean RV difference of 0.07, (95% CI: 0.01 to 0.13) than the SF-36 scales.

THE TRANSLATION AND LINGUISTIC VALIDATION OF THE EQ-5D VISUAL ANALOGUE SCALE (VAS)

Clough LA, Ashcroft-Jones A1, Furtado T, Wild D1,滚动2, (2010); 3)滚动1,滚动2,滚动3

OBJECTIVES: The EQ-5D has been translated into many languages. The Euroqol group have recently altered and clarified the VAS scale. The objective of this study was to produce translations that are conceptually equivalent to the original and to other language versions, ensuring the relevance of the translations within the target cultural context. METHODS: A standard methodology was employed: 1) forward and back translation, review and developer review; or an in-country reviewer and developer review; linguistic validation interviews with 8 subjects, a mix of healthy people and patients, a second developer review and 2 proofreadings. RESULTS: The translation process highlighted numerous cultural and linguistic issues including: 1) Cognitive interviews showed that there was no clear Dutch word for scale, so an explanation likening the scale to a thermometer in the previous 3L VAS was necessary; 2) In some cultures ‘mark an X on the scale’ was difficult to render, and had to be amended by using alternative verb formations and formatting; 3) Though the new VAS mentions only ‘health’, in some languages, it was necessary to use ‘health state’ to avoid confusion, e.g. in Czech “health” alone means “good health.”; 4) In some languages the concepts of “health” and “health state” had different temporal associations. In Korean “health” referred to a longer period of time, so “health today” was used instead of “health state today”; 5) Russian patients understood “health state” as the evaluation given by a doctor or test results, therefore “in your opinion” was added. CONCLUSIONS: The EQ-5D VAS has been translated and linguistically validated using a rigorous translation process. A number of cultural and linguistic issues became apparent and were resolved. The measure is now appropriate for use in multinational trials.

PATIENT-REPORTED OUTCOMES IN PRODUCT DEVELOPMENT GUIDANCE

Mordin M1, Lewis S2, Gnansounou A3, Demuro-Mercon C4, Copley-Merriman K5, Fehmi S6
1RTI Health Solutions, Ann Arbor, MI, USA,2RTI Health Solutions, Research Triangle Park, NC, USA, 3Novartis Pharma, East Hanover, NJ, USA

OBJECTIVES: Patient-reported outcomes (PRO) have received increasing attention from regulatory agencies regarding the demonstration of good clinical practice and new claims. However, some disease areas and/or regulatory bodies necessitate the use of PRO data to substantiate product efficacy for securing approval. Therefore, the research objective was to determine how many of the final product development guidance documents from EMEA and FDA contain guidance on PRO. METHODS: Final guidance documents from the EMEA and FDA were reviewed for mention of PRO. RESULTS: Of the 134 final guidance documents reviewed (EMEA = 81, FDA = 53), 32 mention PRO (EMEA n = 39; FDA n = 13). Within EMEA, PRO is
indicated as primary (n = 5) or secondary (n = 22, of which 4 are secondary and/or exploratory) or both (n = 12). The majority of PRO statements are characterized as signs and symptom measures followed by HRQoL measures. Within FDA, 5 required PRO and 8 suggest use of PRO. The majority of PRO statements are characterized as signs and symptom measures, followed by measures of functional feeling. CONCLUSIONS: PRO data in many disease areas are viewed by regulatory agencies as supportive evidence of the primary endpoint. PRO data are essential in the support of product submissions to regulatory stakeholders, especially within EMEA.

THE IMPACT OF A HOST COUNTRY’S CULTURE ON IMMIGRANT LANGUAGE
Herzberg P1, Flawright R1, Bradley C1, Weatherley BD1, Geldisova M1, Unemori ET1, Conway K1

1MAPI Institute, Lyon, France; Royal Holloway University of London, Surrey, UK
2Momentum Research Inc, Durham, NC, USA; 3Novartis Pharma AG, Basel, Switzerland
4Corhosa, San Mattias, CA, USA

OBJECTIVES: To facilitate international comparison of data, PRO translations must be conceptually equivalent to the original and culturally relevant to the target country. To assess the relevance of conducting a multi-step process on a PRO translation with the aim of using it on an immigrant population speaking that language in a different culture, we investigated the presence and nature of differences between the 2 language versions thus obtained. METHODS: Three translations were compared before and after adaptation to the context of a host country: 1) the Turkish and German Turkish version of the Diabetes Treatment Satisfaction Questionnaire (DTSQ); 2) the original Gujarati and UK Gujarati version of the Subject Self Report on Symptoms Worksheet (SSRSW); and 3) the Chinese Mandarin and US Mandarin version of the National Eye Institute Visual Functioning Questionnaire (NEI-VFQ-25). RESULTS: Six of the eight translated DTSQ were modified following cognate definitions and the use of Turkish speakers in Germany. The Turkish population in Germany tends to use more old-fashioned wording which doesn’t reflect the original language’s recent evolution. All four items in the Gujarati SSRSW needed changing when adapting it to a UK context. Some initially translated wording was reverted back to English, or substituated with transliterated English terms. In the Mandari NEI-VFQ-25, out of 29 items, 11 were modified when adapting it for the USA. The language used in the initial translation was considered too basic for the target population in the USA, which tends to have a higher level of education. CONCLUSIONS: Such a scoring system that separates common outcomes between choices may occasion preferences unaffected by treatment choice. Two questions had a “standard frame”, indicative of commonly asked questions in the literature. The remaining two questions had a “sure thing” frame, in which common outcomes between the two treatments were ranked randomly. Frame order was randomized for each of the four question pairs. A very small, but not statistically significant, difference in QALY maximization (indicated by switched preference) in the “sure-thing” frame, the proportion was 0.08 (15/183). The difference between groups was statistically significant (p < 0.001) CONCLUSIONS: The most common way of asking for preferences for equality tends to foster aversion to inequality, which does not support QALY maximization. In contrast, a frame that separates common outcomes between choices may occasion preferences that maximize QALYs. These results have implications for measurement techniques such as the person tradeoff which assumes framing has no effect on preferences for health allocation.

A UNIVERSAL SCORING SYSTEM FOR EQ-SD: A VASTLY SIMPLER SOLUTION
Kod P1

1University of York, York, UK

OBJECTIVES: Country-specific social preference sets have been estimated to support the use of EQ-SD in comparing QALYs for cost-utility analysis. However, these values have limited applicability in non-economic applications since a) they incorporate the state “dead” which is irrelevant in many therapeutic settings, and b) they are based on hypothetical preferences from 3rd parties who may not have any experience of specific EQ-SD health states. This paper reports on the construction of a scoring system for EQ-SD based on self-rated VAS values generated by individuals with current experience of those states. METHODS: EQ-SD data from different UK sources were pooled yielding a total of 23,679 usable observations. The health state defined by each respondent’s self-rated problem level on the 5 EQ-SD dimensions was determined, yielding a total of 139 unique EQ-SD health states. The mean VAS rating was computed for each of these states. 0/1 dummy variables were defined for each of the EQ-SD dimensions and an OLS regression analysis was performed with the mean self-rated VAS rating as the dependent variable. RESULTS: The model fitted the mean VAS ratings data very well (r² = 0.985) when forced through the origin. All deciles within dimension were monotonic and internally consistent. Residuals were 5 points or lower when observed and estimated values were compared. Estimated values for all EQ-SD health states were computed so that full health (11111) has a value of 100 and worst possible health (33333) has a value of 0. CONCLUSIONS: This methodology contrasts markedly with the more complex requirements of utility estimation and the support of product submissions to regulatory stakeholders, especially within EMEA.

EVALUATING TREATMENT SATISFACTION ENDPOINT EVIDENCE FOR EMEA REGULATORY APPROVALS
Nixon A1, Quadri N, Gallop K

1Oxford Osteoarthritis Ltd, Oxford, Oxon, UK

OBJECTIVES: To document the extent to which treatment satisfaction evidence is provided in support of EMEA regulatory approvals and to evaluate the quality of evidence provided in support of treatment satisfaction claims.METHODS: A review of EMEA published reports for all drugs approved since a centralised process was established in 1995 was undertaken. Specifically the Scientific Discussion/Public Assessment Reports were reviewed for evaluations of patient-reported treatment satisfaction. The wording and types of PROs contained within approved product labels were examined in order to establish the nature and extent of previous successful claims for treatment satisfaction RESULTS: A total of 588 currently authorised medicinal product approvals were reviewed, 26 made reference to ‘satisfaction’ or ‘satisfied’ but 9 were excluded for not focusing on patient-reported treatment satisfaction thus 17 medicinal products were identified as having a direct reference to evaluating patient-reported treatment satisfaction. The 17 approvals ranged from 1998 to July 2008, and were distributed across a broad range of pharma-therapeutic groups with a cluster of approvals for ‘insulin analogues for injection, long lasting’ (n = 4): 10/17 approvals provided limited reference to the way in which treatment satisfaction was evaluated. For example, e.g. reference to a total satisfaction score without any further detail. Satisfactory measured treatment satisfaction using a VAS; 5/17 referenced a specific treatment satisfaction measure. 5/17 provided treatment satisfaction of results, yet only two of these gave any details on the way in which treatment satisfaction was measured.