a pilot project to collect this data and make it available for evaluation and planning processes in a standardized language according to CAP (Catalog for ambulatory procedures). METHODS: The data will be transferred into a new database in two data streams to ensure data protection. Data about the patient containing sex, age and other characteristics will be sent in a pseudonymized way in one data stream. Another will contain data about the procedures according to CAP and information about contract physicians, outpatient clinics and ambulatories. RESULTS: The new database will offer information about what until now has been more or less a black spot. It will give information about procedures performed in the ambulatory- and outpatient sector for all stakeholders participating in this pilot project. CONCLUSIONS: Data about outpatient clinics and ambulatories have not been made accessible in one database for all participating stakeholders in a standardized language until now. The initiated pilot project and the database created therewith offer an opportunity to cover this lack of information.

CONCEPTUAL PAPERS & RESEARCH ON METHODS – Modeling Methods

PMCl2 RULING OUT EXTENDLY DOMINATED OPTIONS USING AN ICER MATRIX
O’Day S, Messner B, Bramley T
Xcenda, Palm Harbor, FL, USA
BACKGROUND: Incremental cost-effectiveness ratios (ICERs) represent the cost per unit of effectiveness of switching to a more costly and more effective option. In reporting results for cost-effectiveness (CE) analyses, options that are strictly dominated are ruled out and no ICERs should be reported. Additionally, some options may be ruled out because of extended dominance (i.e., there is a linear combination of two options that dominates an option not otherwise excluded by strict dominance). In order to plot the CE efficiency frontier both strictly and extendedly dominated options must be excluded. Calculating strict dominance (e.g., in Excel) is straightforward. However calculating extended dominance is more complex. METHODS: We present a method to exclude extendedly dominated options using an ICER matrix. To form an ICER matrix all options are ranked ordered by cost. For a CE analysis with N options, the ICER matrix is an N x N-1 sized table, where the first column represents the ICER from the least costly option to each more costly option, the second column represents the ICER from the second least costly option to each more costly option, etc., Negative ICERs, representing strictly dominated options, are excluded from the table. Extended dominance is established by calculating whether the ICER for a non-strictly dominated option is greater than the ICER for at least one more costly option. If so, the option is ruled out by extended dominance, otherwise not. We show how to perform the required calculations in Excel and how to graphically plot the CE efficiency frontier once all dominated and extendedly dominated options have been excluded. CONCLUSIONS: Strictly dominated and extendedly dominated options must be ruled out in order to plot the CE efficiency frontier. The ICER matrix is a systematic method to rule out strictly and extendedly dominated options.

THE COST-EFFECTIVENESS SENSITIVITY CURVE: QUANTIFYING THE EFFECT OF INDIVIDUAL PARAMETER UNCERTAINTY IN A PROBABILISTIC MODEL
O’Day S, Messner B, Bramley T
Xcenda, Palm Harbor, FL, USA
BACKGROUND: The cost-effectiveness acceptability curve (CEAC) graphically depicts the joint uncertainty in a probabilistic model by transforming the incremental cost-effectiveness ratio (ICER) to a net-benefit framework to represent the probability that a strategy is cost-effective over a range of willingness-to-pay (WTP) thresholds. By depicting the joint uncertainty in a probabilistic model, the CEAC allows decision makers to identify the preferred strategy is cost-effective over a range of willingness-to-pay (WTP) thresholds. By depicting the joint uncertainty in a probabilistic model, the CEAC allows decision makers to identify the preferred

PMCl5 SYNCHRONIZATION OF RANDOM NUMBER STREAMS GREATLY ENHANCES EFFICIENCY OF PROBABILISTIC MODELS
Smedley L, Klein RV
Medical Decision Modeling Inc., Indianapolis, IN, USA
The inclusion of probabilistic components in health care models requires the implementation of random number sampling. Many microsimulation models use a simple random number generator without concern for its properties. This limits the modeler’s ability to use the identical sets of random number streams to assess treatment strategies or among comparators. Synchronization of model results between treatment arms within a model run is difficult unless separate random streams are used for each source of variation. Synchronization of random number streams requires maintaining arrays of starting and current random number seeds. If each source of variation (e.g., times or probabilities of death and major events) has its own stream then the simulation of identical patients in all treatment arms is possible. By resetting the seed for each arm, the model results are impacted only by differences in model input parameter estimates (for first-order analyses) or differences in the specified distributions of sampled model inputs (for second-order analyses). The impact of random number sampling is thus maximally correlated between treatments, and the differences in the occurrences of events common to both simulations, e.g., natural death, are not artificially inflated due to random number sampling. Given a sufficient number of replications, stochastic models usually produce stable results. This is because any further increase in the number of modeled replications will have minimal impact on the average-based model results. If the modeled result is the ratio of differences, such as an incremental cost-effectiveness ratio (ICER), small differences in the denominator often drive the estimate, requiring a large number of replications. The allocation of common events, particularly natural death, to designated random streams minimizes the impact of random sampling on the model results. The number of model replications (and thus execution time) needed to produce stable ICERs may be reduced by as much as 90%.

PMCl6 REVIEW OF COST EFFECTIVENESS STUDIES OF HIGH BUDGET DRUGS
Aggarwal SS, White N, Stevens CA1
1PAREXEL Consulting, Bethesda, MD, USA, 2PAREXEL Consulting, Centreville, VA, USA, 3PAREXEL Consulting, Waltham, MA, USA
OBJECTIVES: The recently made coverage decisions by UK’s NICE, Scotland’s SMC and the allocation of $1.8 Billion for cost effectiveness research by the United States, are strong indicators of trends in pricing and reimbursement that are likely to be observed in the future. To gain an additional insight into these trends, we analyzed the cost effectiveness studies for the top ten highest selling drugs (~$800-95B worldwide sales). METHODS: The Top 10 drugs were selected based on their worldwide sales. For this analysis, we segmented these drugs into categories as primary care, specialty care, small molecules, biologics, therapy areas and availability of generic alternatives. We analyzed the cost effectiveness studies that were published in peer-reviewed journals. Search was conducted using generic common terms, i.e., there is a list of names of the drugs and the phrase “cost effectiveness” in abstract of the published study. RESULTS: During 2003–2008, the number of published studies on “cost effectiveness” have increased by more than 30%. Almost half of the published studies belong to—Remicade, Plavix and Enbrel. There is a large
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 etravirine 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 DIFFERENTIATIVE AND EVALUATIVE PROPERTIES OF THE SF-36 AND THE SF-6D INDEX**

Mutchi A1, Braizer J2, Walters S1

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 of discriminative and evaluative properties. METHODS: The study used relative validity (RV); a ratio of two F statistics, and standardized response mean (SRM) test to evaluate sensitivity and response- ness of the SF-36 scales and SF-6D index. An RV of 1 reflected the most sensitive response scale and the smaller the RV the less sensitive the scale would be. Cohen's criterion for interpreting effect sizes was used to interpret the SRM. The data used were 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. 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 were 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 SRM differences between the SF-36 scales and the SF-6D index in any condition studied. Comparisons 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; 0.06 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 0.15) than the SF-6D 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 the SF-36 index in any condition studied. Comparisons 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; 0.06 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 0.15) than the SF-6D 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 either instrument.

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

Clough LA, Ashcroft-Jones A1, Furtado T, Wild D, Cohen K, Connolly J2

1RTI Health Solutions, Ann Arbor, MI, USA, 2RTI Health Solutions, Research Triangle Park, NC, USA

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 culture. METHODS: A standard methodology was employed: 1 forward and 1 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” had to be expressed as “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 S1, Gnanaakathy A2, Demura-Maron C1, Coffer-Merriman K1, Fehnel S2

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 their role in product development 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 available from EMEA and FDA for clinical and medical research indicate PRO as a mandatory component of efficacy. METHODS: A formal review was conducted of the EMEA and FDA guidance documents that fell under the following categories were excluded: Clinical Pharmacology and Pharmacokinetics, Blood and Blood Forming Organs, Blood products (including plasma derivatives), and Herbals. Included in FDA Guidance review were those listed under the “Clinical/Medical” heading. The following data were abstracted from each guidance document: guidance number, name, issue date, disease area, body system classification, PRO requirement, PRO endpoint hierarchy, and a summary of the PRO language used. 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

**DIMENSIONS CHARACTERIZING GOOD HEALTH BY CHINESE IN CHINA**

Li M1, Bao Z1, Zhou J, Lan N2

1China Pharmaceutical University, Nanjing, Jiangsu, China, The First Affiliated Hospital of College of Medicine, Zhejiang University, Hangzhou, Zhejiang, China, 2National University of Singapore, Singapore, Singapore

OBJECTIVES: Health-related quality of life (HRQoL) instruments used in China are mainly from Western countries. Such instruments may not cover all the important health dimensions relevant to Chinese people as health is a culture-specific concept. Here, we conduct a survey to eliciting characteristics and life domains of good health. RESULTS: Fourteen health dimensions were identified. The 5 most frequently alluded dimensions were: mood (35.5%), absence of disease (33.3%), mobility (25.1%), ability to work (22.4%), and eating (17.5%). Other dimensions included vitality, pain or discomfort, physical fitness, sleep, freedom, self-care, social relationship, enjoyment, and cognition. More proportion of healthy persons than patients quoted mood and self-care as dimensions of health while more patients emphasized ability to work. Males regarded eating as a health dimension more often than females while females quoted self-care and social relationship more frequently than males. With regard to age, older persons valued ability to work more than younger people while more younger people thought absence of disease important. CONCLUSIONS: The study provides useful information for assessing the adequacy of HRQoL instruments developed in Western countries for the Chinese population in China.