

Descriptions of Concurrent Paper Sessions

Session 1: Modern Measurement Theory and Applications

ITEM BANKING AND COMPUTERIZED ADAPTIVE TESTING IN HEALTH OUTCOMES ASSESSMENT

Chih-Hung Chang, PhD*, Assistant Professor, Northwestern University, Chicago, IL

In the management of chronic diseases, the increased interest in patient-reported outcomes (PROs) assessment has been encouraging. The unmet need is a clinically meaningful, psychometrically robust, user-friendly, dynamic and real-time PROs assessment system. Recent advances in modern test theory (i.e., item response theory - IRT) and computer technology make it possible to deliver PROs questions to arrive at an accurate estimate of patient health status with fewer targeted questions. This presentation will address issues related to item bank and computerized adaptive testing development, including 1) methodology to link or equate commonly used PROs instruments; 2) methods to discern and model the structure of higher-order dimensions of self-reported health; 3) steps and challenges to establish an item bank; and 4) technical considerations of developing a computerized adaptive testing (CAT) platform for real-time health status assessment and reporting.

CHOOSING AMONG ITEM RESPONSE THEORY MODELS

Ronald K. Hambleton, PhD*, Professor and Co-Director, Center for Educational Assessment, University of Massachusetts, Amherst, MA

Item response models have been applied to educational tests and credentialing exams for nearly 30 years, and are receiving wide use today in the Quality of Life research area. These applications are wide-spread because of shortcomings with classical approaches to instrument development (e.g., item statistics that are sample dependent, respondent scores that are test dependent) as well as the attractiveness of many features of item response theory (IRT) models that include (1) item parameter invariance (over samples of respondents), (2) ability parameter invariance (over samples of questions), (3) flexibility in instrument development and score equating, (4) unique estimates of measurement error for respondents and (5) enhanced score reporting options. IRT models come in many varieties (over a 100) to handle (1) unidimensional as well as multidimensional data; (2) binary, polytomous, and continuous response data; and (3) ordered as well as unordered categorical responses. The purposes of this presentation will be to (1) provide a comprehensive review of several of the popular IRT models, and (2) offer criteria that can be used in choosing IRT models for particular applications.

COMPARING STANDARD PSYCHOMETRIC AND RASCH APPROACHES USING THE CES-D

James Stansbury, PhD, Rehabilitation Outcomes Research Center, Department of Veterans Affairs, Gainesville, FL, L Douglas Ried, PhD, Rehabilitation Outcomes Research Center, Dept. Veterans Affairs/College of Pharmacy, Gainesville, FL, Craig R. Velozo, PhD, Rehabilitation Outcomes Research Center/Occupational, Dept. Veterans Affairs/Public Health & Health Professions, Gainesville, FL

Objectives: The goal of this study is to compare standard psychometric and Rasch approaches for measuring depressive symptoms using a validated instrument, the Center for Epidemiological Studies - Depression (CES-D). **Methods:** 4,192 older persons with hypertension were mailed a survey containing the CES-D. Usable responses were obtained from 2,455 community-dwelling older persons. Respondent's age, gender, and health status were similar to the population. The Rasch approach is compared to standard item and confirmatory factor analysis techniques. **Results:** CES-D items answered by these older persons demonstrate psychometric properties comparable to published work. Internal consistency for the sample is good ($\alpha = 0.87$). Confirmatory factor analysis yields a four-factor structure similar to previous work, including somatic-retarded activity, depressed affect, positive affect and interpersonal relationship factors. Rasch analysis suggests that the scaling properties of the instrument can be addressed more parsimoniously, reducing nine items to a three-point response with 11 others adequately scored with dichotomous responses. The preliminary analysis suggests the adequacy of a 15-item instrument. **Conclusions:** Rasch analysis shows a hierarchy of specific items that are suited to predicting higher levels of depression, rather than a total score. The Rasch approach suggests that fewer items may be necessary to enhance the utility of the CES-D for depression screening. **Implications:** An optimized depression screening instrument can reduce patient burden and enhance identification of patients at risk of depression and its complications.

**invited presenter*

USING ITEM RESPONSE THEORY TO IMPROVE THE MEASUREMENT OF PHYSICAL FUNCTION: A STUDY IN RHEUMATOID ARTHRITIS

Marie Martin, PhD, Mark Kosinski, MA, John E. Ware, PhD, QualityMetric, Inc., Lincoln, RI, Tracy Li, PhD, Immunology, Bristol-Myers-Squibb Company, Inc., Princeton, NJ, Rhys Williams, PhD, Global Epidemiology & Outcomes Research, Europe, Bristol-Myers-Squibb Company, Inc., Brussels, Belgium, Ross Maclean, MD, Global Epidemiology & Outcomes Research, Bristol-Myers-Squibb Company, Inc., Princeton, NJ, Jakob B. Bjorner, MD, PhD, QualityMetric, Inc., Lincoln, RI

Introduction. Short form questionnaires may fail to detect treatment response in clinical trials because of low precision, floor or ceiling problems. Item Response Theory (IRT) methods were used to examine the properties of two short-form measures of physical function in the context of a clinical trial of rheumatoid arthritis (RA) and develop a combined measure with an expanded range and enhanced measurement precision. **Methods.** Data were from a 12-month, double-blind, multi-center study of 339 RA patients on a background of methotrexate randomized to Abatacept (CTLA4Ig) or placebo. Modified Health Assessment Questionnaire (MHAQ) and SF-36 (with its Physical Functioning scale, PF10) were administered at baseline, 3, 6 and 12 months. Factor analysis for categorical data and IRT methods were used to examine the measurement properties and compute IRT-based scores. Analyses of variance were used to assess sensitivity to changes in disease severity and treatment response. Relative validity coefficients were used to compare the measures. **Results.** Factor analyses suggested either an overall physical function scale or separate scales for upper and lower extremity function. A Rasch IRT model fit the data. IRT scores based on all items lowered the floor and raised the ceiling in the measure of physical function and increased measurement precision. The overall IRT scale was more precise and responsive than the two separate IRT scales and the two original scales: 30% (MHAQ) and 50% (PF10) more efficient in discriminating among ACR groups and 12% (PF10) to 25% (MHAQ) more efficient in discriminating among treatment groups. **Conclusion.** IRT methodology served to estimate a combined score for physical functioning with increased measurement precision and fewer floor/ceiling problems. This measure affords greater power to detect changes in disease activity and treatment response.

MULTIDIMENSIONAL COMPUTERIZED ADAPTIVE TESTING OF THE EORTC QLQ-C30: BASIC DEVELOPMENTS AND EVALUATIONS

Mogens Groenvold, MD, Morten A. Petersen, MSc, Palliative Medicine, Bispebjerg Hospital, Copenhagen, Denmark, Neil K. Aaronson, PhD, Psychosocial Research and Epidemiology, Netherlands Cancer Institute, Amsterdam, The Netherlands, Peter M. Fayers, PhD, Public Health, Aberdeen University Medical School, Aberdeen, Scotland, UK, Mirjam A. Sprangers, PhD, Medical Psychology, Academic Medical Center, Amsterdam, The Netherlands, Jakob B. Bjorner, MD, QualityMetric, Inc., Lincoln, RI

BACKGROUND. Many quality of life instruments are multidimensional and cover several highly correlated domains, each measured by a multi-item-scale. Nevertheless, using standard, unidimensional computerized adaptive testing (CAT) each domain (dimension) is measured by a separate CAT procedure. This seems inefficient. Multidimensional CAT utilizes the correlations between the domains when estimating the domain scores. This may improve measurement efficiency and may thereby reduce the response burden for the patients and reduce the number of items needed in the item pool. However, developing a large, well-functioning multidimensional CAT environment is very demanding and time-consuming. **METHODS.** We initiated a project to evaluate the possible advantages of using multidimensional CAT to measure three domains from the EORTC QLQ-C30, a widely used multidimensional quality of life questionnaire for cancer patients. We used the 12 items from the physical functioning, emotional functioning, and fatigue scales. The evaluations were based on a database with 2,958 European patients. **RESULTS.** Using multidimensional CAT it was possible to reduce the number of items administered to 5-7 items with no or little loss of measurement precision compared to using all 12 items. However, the project also revealed that the estimation of the model underlying the multidimensional CAT and the conceptual aspects require further investigation. **CONCLUSIONS.** Our evaluations to date indicate that multidimensional CAT may significantly improve the measurement precision and efficiency of a quality of life instrument. Further investigations of the use of multidimensional CAT are encouraged.

EVALUATION OF A COMPUTER ADAPTIVE TEST FOR DEPRESSION

M Rose, MD, PhD, H Fliege, PhD, OB Walter, MD, R Kocalevent, J Becker, P Georgiewa, Psychosomatic Medicine, Charité University Medicine, Berlin, Germany, J Bjorner, QualityMetric, Inc., Lincoln, RI, BF Klapp, Psychosomatic Medicine, Charité University Medicine, Berlin, Germany

Depression is one of the most relevant mental health problems and measurement of depressive symptoms becomes more and more important. Computer Adaptive Tests (CAT) promise to provide improved measurements, but there was no application developed for “Depression” so far.- To develop the instrument we used the data from 3,270 psychosomatic patients which answered 11 validated questionnaires undergoing their routine diagnostic examinations. 1528 different patients were used for the evaluation of the itempool in a simulation study. 114 patients and 201 healthy persons answered the D-CAT in its final form. - Three reviewers rated 144 of available 320 items as being representative of the depression construct. All items underwent six steps of analysis to check for the assumptions of unidimensionality, local independence and item discrimination (CFA, residual correlation, trace lines, DIF, simulations). The validation sample was diagnosed using a structured interview (CIDI). Items surveying the more body-related DSM-IV criteria (loss of appetite, weight loss...) did not fulfill the assumptions required for IRT-analysis, while 64 items covering depressive mood, concentration, self-esteem, and suicidal thoughts could be used to apply a General Partial Credit Model. The D-CAT scores were estimated using an EAP-algorithm with a predefined standard error of $<.32$. The latent trait can be computed out of approx. 6 items within 1.7 ± 1.1 minutes. The D-CAT-scores correlated with scores of all available items ($r=.95$) and with the CES-D depression score ($r=.85$), showing a better discrimination of the D-CAT at the low and high ends of the latent trait continuum. D-CAT-scores were the highest for patients with depressive disorders, followed by all other diseases and healthy persons ($p<0.001$). - It can be concluded that depressive mood can be measured economically and with high precision using a CAT.

Session 2: Cognitive Sciences and Health Outcomes Assessment

MEASURING HEALTH RELATED QUALITY OF LIFE: DOES PROSPECT THEORY HELP?

Eve Wittenberg, PhD*, MGH Institute for Technology Assessment, Boston, MA

Health related quality of life is measured with many metrics, including preference-based measures and utilities. This presentation will examine the basis of Cumulative Prospect Theory (CPT), a descriptive alternative to Expected Utility Theory, proposed to explain individuals' preference for decisions made under uncertainty. It will discuss research exploring the application of CPT to decision making in health, and situations in which the theory may be useful.

HOW RELATIONS BETWEEN SELF-REGULATORY ORIENTATIONS AND STRATEGIC ACTIONS INFLUENCE THE QUALITY OF LIFE AS IT UNFOLDS: A THEORY OF REGULATORY FIT

Antonio L. Freitas, PhD*, Dept. of Psychology State University of New York, Stony Brook, NY

Along with people's valuations of desired health outcomes, their valuations of their experiences as they work toward those outcomes impact substantially their motivation and persistence, thus influencing their likelihood of successful outcome realization. Understanding health-related quality of life, then, requires understanding the psychological processes that influence how people come to value and enjoy the health-related activities in which they engage. This talk will address these issues from the standpoint of “regulatory fit,” a conceptual model of the motivational implications of inter-relationships between self-regulatory states and actions. The model suggests that the same action will be experienced differentially positively as a function of the degree to which its strategic inclination is concordant with one's more general self-regulatory orientation. For example, people oriented toward accomplishment experience eagerness-related actions more favorably than vigilance-related actions, whereas people oriented toward responsibility experience vigilance-related actions more favorably than eagerness-related actions. Implications for improving health-related quality of life will be discussed.

COGNITIVE FACTORS IN MENTAL HEALTH OUTCOME MEASUREMENT

Martha Shumway, PhD, George J. Unick, MSW, Tetine Sentell, PhD, Wynne Bamberg, BA, Psychiatry, University of California, San Francisco, CA

Research on standardized questions in opinion surveys documents that cognitive factors affect measurement quality. Regardless of content, the cognitive complexity of questions is associated with responses and persons with limited cognitive capacities are particularly sensitive to question complexity. This study tested the hypothesis that cognitive factors have similar effects on standardized mental health outcome measures. Data on a 32-item measure (BASIS-32) from 1,324 community mental health clients were examined. Dependent variables were question response and objectively identifiable response effects associated with comprehension difficulties (choosing the first, last or middle response, repeating the previous response, no response). Independent variables were indicators of question cognitive

**invited presenter*

complexity (length, position, readability, vague noun phrases, unfamiliar terms) and indicators of respondent cognitive capacity (age, education, language ability, psychosocial functioning, psychiatric diagnosis). Regardless of content, all 5 indicators of question complexity were associated with responses ($p < .01$) and with the 5 hypothesized response effects ($p < .05$). Respondents with limited cognitive capacities (less education, limited language ability and advanced age) were more sensitive to question complexity. Psychiatric diagnosis was also associated with increased sensitivity to question complexity, particularly a diagnosis of schizophrenia, a psychotic disorder. Cognitive factors appear to have similar effects on standardized mental health outcome measures as they do on opinion surveys. Respondents faced with complex items appear to take cognitive shortcuts by choosing the first, last or middle response or offering no response. Those with limited cognitive capacities appear particularly likely to take such shortcuts. Cognitive factors merit increased attention in measure development and selection because their systematic effects seem sufficient to reduce measurement accuracy and compromise measurement equivalence across patient subgroups.

DEVELOPING A SYMPTOMS DIARY - THE CONTRIBUTION OF CASM TECHNIQUES

Elaine McColl, MSc, Cathy M. Stark, PhD, John Bond, BA, Centre for Health Services Research, University of Newcastle upon Tyne, Newcastle upon Tyne, UK

In developing a symptom diary for use in two trials of the management of chronic constipation in older people, we used two techniques of CASM, namely a structured questionnaire appraisal system and individual cognitive interviews, to identify potential threats to data validity and reliability and thus to refine the diary. The draft diary was designed to collect data, on a daily basis, regarding the presence and severity of symptoms of constipation. The choice of symptoms and the wording of items was based on the Rome II criteria. The diary was formatted as a page-per-day, and most questions could be answered simply by ticking the appropriate box. 16 older adults were invited to complete the draft diary for 2-4 weeks and then to participate in a cognitive interview. A total of 14 interviews were conducted by a single member of the study team (CS), in the respondent's home, generally within 3 days of diary completion. We asked about respondents' experiences of completing the diary, and about their understanding of the terms used. Interviews were audio-taped; the completed diaries, tapes and contemporaneous field notes were reviewed by CS. Alongside this patient-centered approach, two members of the study team (EM and JB) independently completed the structured Questionnaire Appraisal System 99 (QAS-99; Willis and Lessler, 1999). Both CASM techniques identified potential problems with the symptom diary. For some respondents to the cognitive interviews, the term 'bowel movement' raised queries; should the passing of a very small amount of waste be counted? Both the QAS-99 and the cognitive interviews identified potential difficulties in responding to a series of questions on difficulties in passing a stool; there appeared to be an implicit assumption of uniform experience, which did not reflect the reality for many respondents. Our findings led us to refine the diary in a number of ways, for example by including a glossary of the terms used to describe the symptoms. Willis GB and Lessler JT (1999) Questionnaire Appraisal Scheme QAS-99. Rockville, MD.: Research Triangle Institute.

PATIENT-REPORTED INSTRUMENT TO ASSESS THE FUNCTIONAL STATUS OF PATIENTS WITH BIPOLAR DISORDER

Sheri E. Fehnel, PhD, Patient-Reported Outcomes, RTI Health Solutions, Research Triangle Park, NC, Timothy Victor, PhD, Epidemiology, AstraZeneca Pharmaceuticals LP, Wilmington, DE, Ranjani Manjunath, MSPH, Patient-Reported Outcomes, RTI Health Solutions, Research Triangle Park, NC, Lee Sechrest, PhD, Psychology, University of Arizona, Tucson, AZ, Stephen Haynes, PhD, Psychology, University of Hawaii at Manoa, Oahu, HI, Stephen Strakowski, MD, Psychiatry, University of Cincinnati College of Medicine, Cincinnati, OH, Joseph Goldberg, MD, Psychiatry, Zucker Hillside Hospital, Glen Oaks, NY, Joyce Cramer, BS, Psychiatry, Yale University School of Medicine, West Haven, CT, Jamie Mullen, MD, Research & Development, AstraZeneca Pharmaceuticals LP, Wilmington, DE, Trisha Suppes, MD, PhD, Psychiatry, University of Texas Medical Center, Dallas, TX

Bipolar disease (BPD) and its treatment often impair patients' daily functioning. This study was designed to identify a sensitive, psychometrically sound patient-reported instrument that ascertains which treatments maximize the functionality of patients with BPD. Through consultation with key opinion leaders, literature review, and individual in-depth interviews with patients with BPD, a questionnaire was developed using 50 items to address the following domains: cognitive functioning, sleep, role functioning, emotional functioning, energy/vitality, social functioning, personal management, and sexual functioning. The draft questionnaire was tested and revised through 2 iterative sets of cognitive interviews with 19 additional patients in multiple locations. In general, the pretest participants deemed the set of constructs addressed in the questionnaire both comprehensive and representative of their daily functioning. They also reported that the final set of items was easy to understand and to fill in, noting that the 7

point Likert-type response scale seemed optimal; the points on the scale appeared to represent the full spectrum of answer choices, yet participants could easily distinguish between the options. Cognitive testing also resulted in the elimination of 17 items deemed either inessential to the measurement of functional status, applicable only to a subset of patients (eg, family responsibilities), or too similar to other items. The resulting questionnaire addresses all constructs considered central to the functional status of patients with BPD, with 33 items phrased to facilitate patient comprehension and completion. A multisite 600-patient validation study is currently evaluating the psychometric properties of this instrument.

Session 3: Advanced Statistical Analysis I

APPLICATIONS OF GEE FOR HANDLING MISSING DATA IN LONGITUDINAL STUDIES

Joseph W. Hogan, PhD*, Center for Statistical Sciences, Brown University, Providence, RI

Generalized estimating equations can be used to fit generalized linear regression models to correlated data, and have been used extensively for longitudinal data. We give a very brief review of the GEE methodology, review some of the key assumptions as they relate to missing data, and illustrate some recently developed modifications for dealing with missing data. Examples from a smoking cessation study and from an HIV cohort study are used to illustrate key concepts. The focus of this talk is intended to be conceptual.

MISSING DATA IN HRQL STUDIES WITH DROPOUT ASSOCIATED WITH MORBIDITY AND MORTALITY

Diane Fairclough, DrPH*, Colorado Health Outcomes Center, University of Colorado Health Sciences Center, Denver, CO

Missing data presents one of the most significant challenges to the interpretations of studies evaluating the health-related quality of life (HRQOL) of patients on a clinical trial. When data are missing due to morbidity or mortality the impact on the assessment of change over time may be dramatic. The methods of analysis that are possible include assumptions that the missing data are ignorable, pattern mixture models, shared parameter models and selection models. What all of these approaches have in common are that they make either an explicit or implicit imputation of the missing values. I will illustrate how graphical presentation of these imputed values can help us gain a better understanding of the underlying assumptions of these models using data from clinical trials with dropout associated with mortality and morbidity. These approaches have in common that they make either an explicit or implicit imputation of the missing values. I will illustrate how graphical presentation of these imputed values can help us gain a better understanding of the underlying assumptions of these models using data from clinical trials with dropout associated with mortality and morbidity.

THE USE OF BOOTSTRAP METHODS FOR ANALYSING HEALTH-RELATED QUALITY OF LIFE OUTCOMES (PARTICULARLY THE SF-36)

Stephen J. Walters, PhD, Michael J. Campbell, PhD, ScHARR, University of Sheffield, Sheffield, UK

Health-Related Quality of Life (HRQoL) measures are becoming increasingly used in clinical trials as primary outcome measures. Investigators are now asking statisticians for advice on how to analyse studies using HRQoL outcomes. HRQoL outcomes like the SF-36 are usually measured on an ordinal scale. However, most investigators assume that there exists an underlying continuous latent variable that measures HRQoL, and that the actual measured outcomes (the ordered categories), reflect contiguous intervals along this continuum. The ordinal scaling of HRQoL measures means they tend to generate data that have discrete, bounded and skewed distributions. Thus, standard methods of analysis such as the t-test and linear regression that assume Normality and constant variance may not be appropriate. For this reason, non-parametric methods are often used to analyse HRQoL data. The bootstrap is one such computer intensive non-parametric method for analysing data. We used the bootstrap for hypothesis testing and the estimation of standard errors and confidence intervals for parameters, in four datasets (which illustrate the different aspects of study design). We then compared and contrasted the bootstrap with standard methods of analysing HRQoL outcomes as described in Fayers and Machin (2000) and Fairclough (2002). The standard methods included analysis of covariance, summary measures and Generalised Estimating Equations. Overall, in the datasets studied with the SF-36 outcome the use of the bootstrap for analysing HRQoL data appears to produce results similar to conventional statistical methods. Therefore, the results of this work suggest that bootstrap methods are not more appropriate for analysing HRQoL outcome data than standard methods. This result requires confirmation with other HRQoL outcome measures, interventions and populations. er, Wiley. Fairclough D.L. (2002) Design and Analysis of Quality of Life Studies in Clinical Trials. New York, Chapman & Hall.

**invited presenter*

LONGITUDINAL CHANGES IN PERCEIVED ENERGY BEFORE AND AFTER LUNG TRANSPLANTATION: COMPARING RESULTS OF COMPLETE CASE, CROSS-SECTIONAL, AND MULTI-LEVEL ANALYSIS

Karin M. Vermeulen, MSc, Wendy J. Post, PhD, Office for Medical Technology Assessment, Gerard H. Koëter, PhD, Pulmonary Diseases, Elisabeth M. TenVergert, PhD, Office for Medical Technology Assessment, Groningen University Hospital, Groningen, The Netherlands

Introduction: Analyses of HRQL are often hampered by patients who drop out, due to poor health or death. To assess changes over time, longitudinal analyses of complete cases are often performed, which may cause bias. Aim of our study was to compare the results of complete case analysis, cross-sectional analysis, and the results of a multi-level model in which all patients were included until they dropped out. **Methods:** 473 Patients completed one or more self administered HRQL questionnaires during the waiting period and/or after lung transplantation. HRQL was measured at specific points in follow-up. For illustrative purposes we focussed on the dimension energy of the Nottingham Health Profile (NHP). Changes in NHP-energy scores were analyzed using the three strategies. **Results:** All three types of analysis showed that NHP-energy scores improved significantly after LgTX and remained more or less constant over time. In both waiting list and transplanted patients the complete case analysis showed more favorable scores compared to the cross-sectional, and the multi-level analysis. In the multi-level model the predictors age, gender, and Bronchiolitis Obliterans Syndrome appeared to have a significant influence on NHP-energy scores. **Conclusions:** Our study showed that complete case analysis overestimated the scores compared to the other two strategies. Furthermore, in cross-sectional analyses different groups of patients are analyzed at different time points. Therefore in cross-sectional analysis no changes over time can be assessed. Since in multi-level modeling all available data are used, this is a feasible approach to analyze longitudinal changes in studies where missingness of data is a problem.

A UNIFIED FRAMEWORK FOR SCORING AND MISSING DATA ESTIMATION FOR THE SF-36, SF-12, AND SF-8

Renee N. Saris-Baglana, PhD, Jakob B. Bjorner, MD, PhD, Mark Kosinski, MA, John E. Ware, PhD, Measurement Science, QualityMetric, Inc., Lincoln, RI

To enhance the usefulness of the SF tools, we developed scoring methods that enabled comparison of scale scores across tools (scale calibration) and missing data estimation (MDE). Norm based scoring (NBS) and Item Response Theory (IRT) methods were used to achieve scale calibration. IRT and regression methods were used to develop improved methods for MDE. Algorithms were developed from the 1990 and 1998 National Surveys of Functional Health Status (NSFHS) and were evaluated in the NSFHS and the Medical Outcomes Study (MOS). MDE was evaluated by introducing missing data and comparing estimated and original scores. We saw acceptable agreement between the NBS of different SF tools and versions and reached identical conclusions for tests of group differences in nearly all cases. MDE was possible: 1) for a scale score if a respondent had answered at least one item representing the construct; 2) for the PCS score if a respondent had a score on the Physical Functioning (PF) scale and at least 6 other scales; and 3) for the MCS score if a respondent had a score on the Mental Health (MH) scale and at least 6 other scales. The agreement between estimated and actual scale and summary scores was very high ($r=.89-.97$ for scales; $r=.94-.99$ for summaries) and the mean estimated scale and summary scores never differed by more than 1.9 points from original scores. Our MDE approach recovered scores for more than 85% of respondents with partially completed survey responses. These results suggest that our scoring algorithms permit comparison of scores across SF tools and versions and that scores can be estimated without bias for many respondents who otherwise would have missing scores. To enhance the standardization of these approaches, we have integrated them in a scoring software package.

QUALITY OF LIFE (QOL) RESULTS OF A RANDOMIZED STUDY OF INTRAVENOUS (IV) PACLITAXEL AND CISPLATIN VS IV PACLITAXEL, INTRAPERITONEAL (IP) CISPLATIN AND IP PACLITAXEL IN OPTIMAL STAGE III EPITHELIAL OVARIAN CANCER (OC): A GYNECOLOGIC ONCOLOGY GROUP TRIAL

Lari B. Wenzel, PhD, College of Medicine, University of California Irvine, Irvine, CA, Helen Huang, MS, GOG Statistical Office, Roswell Park, Buffalo, NY, Deborah Armstrong, MD, Oncology, Johns Hopkins Kimmel Cancer Center, Baltimore, MD, Joan Walker, MD, Gynecologic Oncology, University of Oklahoma, Oklahoma City, OK, David Cella, PhD, Center on Outcomes Research and Education, Northwestern University, Evanston, IL

A study examining IP therapy in optimal stage III OC patients suggested improvement in relative risk of recurrence for patients in that arm (ASCO, 2002). An important secondary outcome was assessment of QOL. 415 patients were randomly allocated to receive Regimen 1 (R1): IV paclitaxel 135mg/m²/24h and IV cisplatin 75mg/m² or Regimen 2 (R2): IV paclitaxel 135mg/m²/24h followed by IP cisplatin (100mg/m²), plus IP paclitaxel 60mg/m² on day 8, with cycles repeated every 21 days x 6. Using the FACT-O, GOG-NTX, and pain measure (PAIN), QOL was assessed: prior to randomization (baseline), prior to cycle 4, and 3-6 weeks and 12 months post-treatment. Multivariate analysis was used to compare baseline FACT-O scores with NTX and PAIN subscales. Follow-up assessments were analyzed using a Linear Mixed Model with unstructured covariance and baseline scores, age and performance status as covariates. Mean FACT-O and PAIN scores were significantly different between treatment arms, with R2 patients reporting worse QOL and pain prior to cycle 4 ($p < 0.0001$) and worse QOL 3-6 weeks post-treatment ($p = 0.0035$). Neurotoxicity was also significantly worse in R2 3-6 weeks after completing chemotherapy ($p = 0.0004$), and one year later ($p = 0.0018$). However, there were no significant QOL or PAIN score differences between arms one year post-treatment. Patients who received higher dose IP therapy (R2) experienced more QOL disruption, pain and NTX when compared to those who received more conventional dose IV therapy (R1). However, R2 patients experienced better recurrence-free survival. This trade-off should be considered when discussing treatment options with patients.

Session 4: Qualitative Research Methods

WHAT OUR WORDS SAY ABOUT US: USE OF COMPUTERIZED TEXT ANALYSIS IN RESEARCH AND PRACTICE

James Pennebaker, PhD*, Dept. of Psychology, University of Texas, Austin, TX

The words people use in natural speech or writing conveys a great deal about their social situation, personality, and current psychological states. Much of this information is inherent in function words such as articles, pronouns, and prepositions. By using a computerized text analysis program, we have been able to demonstrate that function words do much better than chance in predicting depression, suicidality, hormone levels, dominance in a relationship, sex, age, social class, etc. Implications for assessing quality of life will be discussed.

AUTOMATED SYSTEMS THAT ANALYZE TEXT AND DISCOURSE: AUTOTUTOR, COH-METRIX, AND QUAID

Arthur C. Graesser, PhD*, Professor, Dept. of Psychology, University of Memphis, Memphis, TN

Recent advances in computational linguistics, cognitive science, and discourse processes make it more feasible for computers to comprehend text at varying degrees of depth and to respond in an adaptive, flexible manner. The presentation will describe three systems that analyze text and discourse, developed in the Institute for Intelligent Systems at the University of Memphis. AutoTutor simulates human tutorial dialog, with an animated agent that holds conversations with students in natural language. Coh-Metrix is a web facility that analyzes texts on over 200 measures of language, cohesion, and text coherence. QUAID (Question Understanding Aid) is a web facility that critiques survey questions on potential problems of unfamiliar terms, vague relative terms, unclear noun-phrases, syntactic complexity, and working memory overload.

EVALUATION OF THE SEIQOL-DW AS A MEASURE OF QOL USING QUALITATIVE ANALYSIS

Hanne Bruhn, MA, Kristina Lauche, PhD, Saskia Teunisse, PhD, School of Psychology, University of Aberdeen, Aberdeen, Scotland, UK, Louise H. Phillips, PhD, School of Psychology, University of Aberdeen, Aberdeen, UK

The SEIQoL-DW consists of a semi-structured interview eliciting the five most important areas to respondents' quality of life (QoL). Respondents nominate five areas (cues) and also describe what it is about each area that makes it important to them and their QoL. Thereafter respondents rate how they are doing in each area on a visual-analogue scale (0-100 mm) and assign importance to areas by giving more important areas a larger part of a pie chart than less important areas. From these data an overall QoL index can be calculated. The popularity of the SEIQoL-DW as an instrument is mainly based on the possibility of calculating an overall score. Little use is made of the personal information or explanation obtained from the participants. We argue that this qualitative information is of methodological interest and clinical value for the respondent's QoL. Both purposes require a stringent qualitative methodology. For this study the SEIQoL-DW interview of 15 elderly women (mean age = 72.8 years, SD = 5.3) was audio recorded for qualitative analysis as part of another study (N = 35). The explanations of the areas were

**invited presenter*

transcribed verbatim and content analysed. The coding also included connotation (positive, negative and neutral), time span (permanent, transient, age specific). The interrater reliability was 80%. The findings allow us to characterise this sample of elderly females and compare their QoL areas with QoL areas in other studies and show how the SEIQoL-DW can be used for deep structure information. The implications for the measurement of QoL and the validation of the SEIQoL-DW will be discussed.

A QUALITY OF LIFE SCALE FOR CHILDREN WITH CEREBRAL PALSY: QUALITY OF LIFE FROM THE PERSPECTIVE OF FAMILIES

Elizabeth B. Waters, DPhil, Centre for Community Child Health, Royal Children's, Murdoch Childrens Research Institute, Parkville, Victoria, Australia, Elise C. Maher, PhD, Centre for Community Child Health, Royal Children's, Murdoch Childrens Research Institute, Parkville, Victoria, Australia, Dinah Reddihough, MD, Roslyn Boyd, PhD, Child Development and Rehabilitation, Royal Children's Hospital, Parkville, Victoria, Australia

An international condition-specific QOL scale for children with cerebral palsy is being developed to evaluate the effectiveness and impact of treatment interventions in clinical trials. The measure's content is based on determinants of QOL, reported by children with cerebral palsy and their parents. Semi-structured interviews were conducted with 26 children with cerebral palsy and their families to identify the major themes of QOL. Parent-proxy interviews were conducted for children aged 4-8 years, and parent-proxy and child-report interviews were conducted for children aged 9-12 years. Data were analysed to provide thematic summaries, and specific items were derived to measure each domain of QOL through consulting with researchers, clinicians and parents. The items aim to measure several domains including physical health, body pain and discomfort, daily living tasks, participation in regular activities, future QOL, emotional wellbeing and self-esteem, family health, financial stability, provision of, and access to services, supportive physical environment, social wellbeing, communication, and interaction with the community. In order to account for the developmental changes in children and to enable assessment of children with physical, intellectual and communication disability, several versions of the scale are developed including two parent-proxy versions (4-8 years; 9-12 years), and a child self-report version (9-12 years). The resulting QOL scale for children with cerebral palsy includes the traditional domains of QOL such as physical, emotional, social, and family health. It also includes more practical domains such as access to care, financial stability, use of equipment, and acceptance into the broader community. Methodological issues relating to scoring and evaluating the child report and parent-proxy questionnaires will be discussed.

A MIXED METHOD APPROACH TO QUALITATIVELY EXAMINING LATINAS' PERCEPTIONS OF GYNECOLOGICAL HEALTH AND CERVICAL CANCER SURVIVORSHIP

Juliet M. McMullin, PhD, Israel DeAlba, MD, MPH, F. Allan Hubbell, MD, MSPH, Lari Wenzel, PhD, Medicine, University of California Irvine, Irvine, CA

The incidence and mortality rates of cervical cancer in Latina women are nearly twice that of non-Latina white women. This disparity persists in the survivorship literature. Although cervical cancer survivorship studies document significant QoL disruption in non-Latina white women, no published studies have targeted Latina cervical cancer survivors. Drawing on a mixed method approach of qualitative and quantitative data, the goals of this paper are twofold: (1) to examine Latinas' understandings of how cervical cancer has disrupted their QoL. (2) to describe the use of consensus analysis as an analytic method that integrates qualitative findings to measure how much participants share a cultural model. A bilingual interviewer conducted qualitative interviews with 30 Latina cervical cancer survivors and 30 Latina age and income matched controls using a semi-structured questionnaire examining survivorship issues. We evaluated the interviews for major themes using qualitative content analysis. Using some of the frequently mentioned QoL issues, we created a set of items that the women stated impacted their gynecological health. These items were then used to measure whether or not the women shared a cultural model of gynecological health issues that are important to their overall QoL. Major themes included the importance of God, prayer and going to church, talking to and getting help from their family, and lifting their spirits. With regard to the cultural model, however, the survivor group did not agree on the gynecological health issues relative importance to their overall QoL. In contrast, the control group did agree on the gynecological health issues relative importance to their overall QoL. These data highlight the importance of qualitatively examining the ways in which cervical cancer disrupts the QoL of Latinas and the potential shifts in expectations from a healthy population of Latinas.

Session 5: Advanced Statistical Analysis II

Sensitivity Analysis for Longitudinal Clinical Trials

Herbert Thijs, PhD*, Center for Statistics, Limburgs University, Diepenbeek, Belgium

In a longitudinal study or experiment, each unit is measured on several occasions. It is not unusual in practice for some sequences of measurement to terminate early for reasons outside the control of the investigator, and any unit so affected is often called a dropout. Furthermore Rubin (1976) and Little and Rubin (1987, Ch. 6) make important distinctions between different missing values processes. A dropout is independent of both unobserved and observed data and random (MAR) if, conditional on the observed data, the dropout is independent of the unobserved measurements; otherwise the dropout process is termed non-random (MNAR). Currently in clinical trials standard methodology used to analyze longitudinal data subject to non-response is mostly based on the MCAR assumption and includes Last Observation Carried Forward (LOCF), Complete Case analysis, etc. This is often done without questioning possible influence of these assumptions on the final results. On the other hand, if a dropout process is random, then a valid analysis, can be obtained through a likelihood-based analysis that ignores the dropout mechanism, provided the parameters describing the measurement process are functionally independent of the parameters describing the dropout process, the so-called parameter distinctness condition. This situation is termed ignorable by Rubin (1976) and Little and Rubin (1987). This leads to considerable simplification in the analysis. In many examples, however, the reasons for dropout are many and varied and it is therefore difficult to justify on a priori grounds the assumption of random dropout. Arguable, in the presence of non-random dropout, a wholly satisfactory analysis of the data is not feasible. One approach is to estimate from the available data the parameters of a model representing a non-random dropout mechanism. It may be difficult to justify the particular choice of dropout model, and it does not necessarily follow that the data contain information on the parameters of the particular model chosen, but where such information exists the fitted model may provide some insight into the nature of the dropout process and of the sensitivity of the analysis to assumptions about this process. This is the route taken by Diggle and Kenward (1994) in the context of continuous longitudinal data; see also Diggle, Liang and Zeger (1994, Ch. 11). Further approaches are proposed by Laird, Lange, and Stram (1982), Wu and Bailey (1988, 1989), Wu and Carroll (1988), and Greenlees, Reece, and Zieschang (1982). An overview of the different modeling approaches is given by Little (1995). Also the case of categorical outcomes has received considerable attention. See for example Baker and Laird (1988), Stasny (1986), Baker, Rosenberger, and Dersimonian (1992), Conaway (1992, 1993) Park and Brown (1994) and Molenberghs, Kenward and Lesaffre (1997). With the volume of literature on non-random missing data increasing, there has been growing concern about the fact that models often rest on strong assumptions and relatively little evidence from the data themselves. This point was already raised by Glynn, Laird and Rubin (1986) who indicate that this is typical for so-called selection models, where the joint distribution of the measurement and missingness processes is factorized into the marginal distribution of the measurement process and the conditional process of the missingness process given the measurements, while it is much less so for a pattern-mixture model (Little 1993, 1994, Hogan and Laird 1997), where the reverse factorization is used. Since the model of Diggle and Kenward (1994) fits within the class of selection models, it is fair to say that it raised, at first, too high expectations. This was made clear by many discussants of their paper. This implies that, for example, formal tests for the null hypothesis of random missingness, while technically possible, should be approached with caution. In response, there is a growing awareness of the need for methods that investigate the sensitivity of the results with respect to the model assumptions. See for example Nordheim (1984), Little (1994), Rubin (1994), Laird (1994), Fitzmaurice, Molenberghs and Lipsitz (1995), and Molenberghs, Goetghebeur and Lipsitz (1998). Still, only a few actual proposals have been made. Moreover, many of these are to be considered as useful but ad hoc approaches. In our view, a more formal approach to sensitivity analyses should be fruitful as well. Taking this into account and based on a case study from the clinical industry we would like to stress the major drawbacks of a simple LOCF analysis or CC analysis. We will show the discrepancies of both the results from these analyses and we will also compare both analyses with a stronger MAR analysis with regard to the results. Furthermore we will expand our investigation of sensitivity towards MAR and MNAR models, combining a measurement model with a dropout model as done by Diggle and Kenward (1994), allowing the investigator to make prior conclusions with regard to the missingness process. At last, we will also apply some advanced sensitivity tools, such as local and global influence as they are introduced by Verbeke et al (2001) and further discussed by Molenberghs et al (2001) and Thijs, Molenberghs and Verbeke (2003). Using these tools may lead to the detection of influential subjects. In addition, it will be shown that

a pattern mixture approach (Thijs et al 2002) is a viable alternative to gain insight in the specific problems regarding missing data.

BAYESIAN ANALYSIS OF HEALTH STATUS AND QUALITY OF LIFE DATA

Dennis Fryback, PhD*, Professor, Dept. of Population Health Studies, University of Wisconsin-Madison, Madison, WI

The past 15 years has seen large advances in applied Bayesian methodology and the wide distribution of the relatively accessible BUGS and WinBUGS software for Bayesian data analysis. In this talk we will illustrate Bayesian analysis capabilities using SF-36 data (including SF-6D utility scoring) from the Beaver Dam Health Outcomes Study along with long term longitudinal follow-up of this cohort.

QUALITY OF LIFE TRAJECTORIES AMONG MASSACHUSETTS ADULTS WITH SUBSTANCE USE DISORDERS

Kevin W. Smith, MA, Annie Zhang, MPH, Mary Jo Larson, PhD, New England Research Institutes, Watertown, MA

OBJECTIVES: The objective of this analysis was to characterize longitudinal changes in quality of life in a group of adults receiving publicly-funded treatment for substance use disorders and to determine the effects of demographic and treatment factors on quality of life trajectories. **METHODS:** Clients were randomly sampled from 13 Massachusetts facilities providing publicly-funded detoxification and outpatient treatment services to adults diagnosed with substance abuse disorders in 1997-1998. A total of 217 clients completed an in-person baseline interview and follow-up telephone interviews 1 year and 3 years later. The two primary domains of quality of life (QOL), physical functioning and emotional well-being, were measured by the SF-12 Physical Component Score (PCS-12) and Mental Component Score (MCS-12). Clients who died during the follow-up interval were retained in the analyses by assigning utility-based component scores for death. Latent growth modeling was used to estimate the effects of five factors—age, gender, detox status, managed care status, and drug treatment services during the second year—on QOL trajectories over time. **RESULTS:** Seventy percent of the clients were recruited from detoxification facilities, and 52% were enrolled in Medicaid behavioral health carve-out managed care plans. The growth models provided excellent fits for both the MCS-12 (RMSEA=.000) and PCS-12 (RMSEA=.077) trajectories. The mean MCS increased from 30.0 to 38.6 at the time of the one year follow-up and remained stable at year 3 (38.7). Clients recruited from detox centers had significantly lower MCS scores at baseline and much higher slopes over time. The mean PCS increased from 43.4 to 44.6 after 1 year but then declined back to 41.6 by the time of the 3-year follow-up. Baseline physical functioning was negatively correlated with client age and managed care status. Managed care and drug treatment during the intervening period had little impact on QOL trajectories. **CONCLUSIONS:** Latent growth models provide a flexible framework for characterizing trajectories of QOL measures.

RASCH-INFORMED CATEGORIZATIONS OF TRANSITION RATING SCALES

Kathleen W. Wyrwich, PhD, Research Methodology & Health Services Research, Stacie M. Metz, MA, Health Services Research, Saint Louis University, St. Louis, MO, Ajit N. Babu, MBBS, Primary Care, St. Louis Veterans Administration Medical Center, St. Louis, MO, Kurt Kroenke, MD, General Internal Medicine, Indiana University, Indianapolis, IN, William M. Tierney, MD, General Internal Medicine, Indiana University School of Medicine, Indianapolis, IN, Fredric D. Wolinsky, MD, Health Management and Policy, University of Iowa, Iowa City, IA

Patient-perceived change in health-related quality of life (HRQoL) domains has repeatedly been classified using the 15-point patient transition rating scale, ranging from -7 (a very great deal worse) to 0 (about the same) to 7 (a very great deal better), since its introduction by Jaeschke et al. in 1989. The traditional change levels of trivial (-1, 0, or 1), minimal (2, 3 or -2, -3), moderate (4, 5 or -4, -5) and large (6, 7 or -6, -7) on this scale, however, have been arbitrarily defined and originally assumed that change related to an improvement was the same as that for a decline. We used Rasch partial credit models of each domain's transition rating and the changes in each domain's items to determine natural cut points for minimal, moderate, and large differences on the transition rating scale when used to assess patient change on the domains of the Chronic Heart Disease Questionnaire (CHQ), Chronic Respiratory Disease Questionnaire (CRQ), Asthma Quality of Life Questionnaire (AQLQ), and the SF-36, Version 2 scales. Our sample included 1662 heart disease, lung disease or asthmatic outpatients attending primary care clinics in the Midwest US. These outpatients completed telephone interviews on one of the disease-specific HRQoL questionnaires (CHQ, CRQ or AQLQ) and the SF-36 every two months over one year of participation, and also provided transition

**invited presenter*

ratings for each HRQoL disease-specific domain and SF-36 scale at each follow-up interview. The Rasch-informed categorizations for minimal, moderate and large changes in the CHQ domains display symmetry between improvements and declines, and the new moderate change level in the CHQ activities and fatigue domains included ratings of 4, 5 and 6. Minimal improvement (and decline) categorizations in most domains and scales continued to only include ratings of 2 and 3 (-2 and -3). Change thresholds for the SF-36 scales differed slightly among the three disease groups (heart disease, lung disease and asthma patients), but reflected a similar trend across patients groups and most HRQoL measures for including only the rating of 7 in the large improvement category

MIXED EFFECTS MODELING OF THE RELATIONSHIP OF CYTOKINES AND SYMPTOM SEVERITY IN PATIENTS DURING FIRST 30 DAYS OF AUTOLOGOUS BMT

Xin S. Wang, MD, Symptom Research, Sergio A. Giralt, MD, Blood and Marrow Transplantation, Carla L. Warneke, MS, Biostatistics, Tito Mendoza, PhD, Karen O. Anderson, PhD, Charles S. Cleeland, PhD, Symptom Research, UT MD Anderson Cancer Center, Houston, TX

Objective: Patients receiving autologous BMT suffer from multiple symptoms at the acute phase of transplantation, including pain, fatigue, poor appetite, and disturbed sleep. To have more understanding of the mechanism of development of these non-specific symptoms, this pilot study prospectively assessed the possible relationship of changes in cytokines and symptom severity during first 30 days of auto-BMT. Methods: The M.D. Anderson Symptom Inventory (MDASI) was the assessment tool used before and 3 times a week during the first 30 days post-transplantation in 18 patients to measure multiple symptom severity and interferences. Multiple cytokines was assayed pre-transplantation, at nadir, and at 30 days of BMT, including IL-1ra, IL-2, IL-6, IL-8, IL-11, IFN, TNF-a. Results: Serum level of IL-6 significantly increased at nadir and back to almost baseline level at 30 days of BMT. In the mixed effect modeling, the MDASI symptom severity component score is the response variable, the linear term for IL-6 is significantly associated with symptom severity component score ($p = 0.001$), as is the quadratic term ($p = 0.009$). We assume a spatial correlation structure for each patient to account for longitudinal measurements that are unequally spaced in one dimension. To this model, we sequentially added the remaining cytokines. IL-2 is the only additional cytokine that contributes significantly to the model ($p = 0.017$). Conclusion: Preliminary results suggested an observed relationship between increased IL-6 and IL-2 and increased symptom severity level during the acute phase of auto-BMT. Longitudinal modeling is a way to track association between self report symptoms and biological variables.

Session 6: Theoretical Models for HRQL Research

TITLE AND DESCRIPTION TBD

Julia Fox Rushby, PhD*

TOWARD AN INTEGRATIVE PSYCHOLOGICAL THEORY OF QUALITY OF LIFE

Joseph Sirgy, PhD*, Professor, Dept. of Marketing, Pamplin College of Business, Virginia Tech, Blacksburg, VA

The presentation will involve a description of an integrative psychological theory of QOL. The theory posits that much of the research on subjective well being can be construed in terms of personal strategies people use to optimize their happiness and life satisfaction. These strategies include bottom-up spillover, top-down spillover, horizontal spillover, balance, e-evaluation, goal selection, and goal implementation.

TESTING THE WILSON AND CLEARY HRQOL CONCEPTUAL MODEL IN PERSONS LIVING WITH AIDS USING STRUCTURAL EQUATION MODELING

Karen H. Sousa, PhD, Nursing, Oi-Man Kwok, MS, Psychology, Arizona State University, Tempe, AZ

For patients with chronic illnesses health-related quality of life (HRQOL) has emerged as a potentially important outcome of health care. However, debate continues about the dimensions of and contributors of HRQOL in relation to their relevance and relative importance, the strengths of their influence, and their dynamic nature. An atheoretical approach to measuring HRQOL as a multidimensional construct potentially results in a laundry list of variables. Wilson and Cleary (1995) proposed a HRQOL conceptual model that included five dimensions: biological variables, symptoms status, functional health status, general health perceptions, and overall quality of life. The purpose of this analysis was to test the Wilson and Cleary HRQOL conceptual model using structural equation modeling. Data from the AIDS Time-Oriented Health Outcome Study were used for this analysis; 83% of the sample was Caucasian

**invited presenter*

and all were male. Indicators associated with biological variables, symptom status, general health perception, and overall quality of life were continuous while the indicators for the functional health status component were dichotomized. To account for the non-normal properties of the indicators from the functional health status component, analyses were conducted using the weight least square estimation with robust standard errors and a mean-adjusted chi-square statistic test in Mplus. The first model tested (N = 917) contained the symptom status, functional health status, general health perception, and overall quality of life dimensions. This model fit the data adequately ($\chi^2(167) = 1496.68, p < .001; CFI = .98; RMSEA = .083$). All paths were significant. The second model tested (N = 395) included the CD4 count representing the biological variable. The results showed that this model also had reasonable fit ($\chi^2(184) = 459.59, p < .001; CFI = .98; RMSEA = .062$). Implications of these results will be discussed. A valid HRQOL conceptual model would help in understanding the relationship among the predictors and help translate its clinical relevance.

THE EFFECT OF MATERIALISM ON EMOTIONAL WELL-BEING AND LIFE SATISFACTION: AN APPLICATION OF MULTIPLE DISCREPANCIES THEORY

James W. Shaw, PharmD, Pharmaceutical Sciences, Ada Leung, BS, Marketing, Stephen J. Coons, PhD, Pharmaceutical Sciences, Melanie Wallendorf, PhD, Marketing, University of Arizona, Tucson, AZ

Studies have found materialism to be associated with depression, anxiety, and reduced life satisfaction. This research sought to establish and test a theory of the relationship between materialism and quality of life. Materialism has alternatively been characterized as a trait or value. Many researchers have reported a direct relationship between materialism and measures of subjective well-being. However, a direct association seems implausible if one defines materialism as a value instead of a trait. We developed a model using multiple discrepancies theory (MDT) to explain why materialistic individuals report poor quality of life. According to this model, materialists evaluate their standard of living poorly due to unreasonably high goals and expectations. This leads to reduced happiness and/or satisfaction with standard of living (i.e., material well-being), resulting in diminished emotional well-being and life satisfaction. A cross-sectional survey was administered to 737 community respondents living in Tucson, Arizona. The sample was stratified by sex, age, and social class. The structural model was tested using a generalized method of moments instrumental variables estimator. Materialism was found to be negatively related to evaluations of standard of living against social references and perceived needs. These evaluations were positively related to material well-being. Material well-being was positively related to emotional well-being and life satisfaction, and the latter two constructs were also positively correlated. However, there continued to be a negative association between materialism and global measures of quality of life even after controlling for a direct relationship with material well-being. Our research indicates that MDT does not fully explain the relationship between materialism and global measures of quality of life. Our findings suggest that other theories, such as value conflict, should be used to supplement judgement theory when attempting to explain the materialism/quality-of-life relationship.

HAS THE MEASURABLE DRIVEN OUT THE IMPORTANT? INDIVIDUALIZED QUALITY OF LIFE

Lena Ring, PhD, Stefan Höfer, PhD, Psychology, Health Services Research Centre, RCSI, Dublin, UK, Hannah McGee, PhD, Psychology, Health Services Research Centre, RCSI, Dublin, Ireland, Anne Hickey, PhD, Ciaran O'Boyle, PhD, Psychology, Health Services Research Centre, RCSI, Dublin, UK

Over the past two decades there has been a growing interest in the assessment of quality of life (QoL) and hundreds of measures now exist. Most of these adopt an operational definition and are not grounded in any psychological or philosophical theory of QoL. Empirical findings show that QoL is highly individual and that people vary in the importance they attach to different life areas. Emphasis on autonomy means that the quality of an individual's life is neither more nor less than what she considers it to be, and is, as such, based on subjective judgment. Individual QoL (IQoL) assessment (the idiographic approach) is reflected in, for example, the cognitive approach to personality (universal dimensions are not assumed) and in the hermeneutic approach to happiness. Methods based on the repertory grid and social judgment theories allow the construction of a person's cognitive space with reference to individually important dimensions. A range of individualised QoL (IQoL) measures have been developed and can be classified according to different levels of individualisation. Calman's gap theory of QoL (1987) can be used as one conceptual base for IQoL measures. Different proposed models of QoL will be discussed in relation to the IQoL approach, e.g., Keyes' & Ryff's well-being model (2002), Seligman's authentic happiness model (2002) and Sirgy's subjective well-being model (2002). Empirical examples showing the individualised nature (domains and their importance) as well as the dynamic nature (adaptation/response shift) of QoL will be presented. It is concluded

that, if QoL is viewed as highly individual, then the measures purporting to capture it should likewise be individualised.

RELATIONSHIP OF INDIVIDUAL QUALITY OF LIFE WITH PSYCHOLOGICAL AND SUBJECTIVE WELL-BEING

Stefan Höfer, PhD, Lena Ring, PhD, Psychology, Royal College of Surgeons in Ireland, Dublin, UK, Hannah McGee, PhD, Psychology, Royal College of Surgeons in Ireland, Dublin, Ireland, Anne Hickey, PhD, Ciaran A. O'Boyle, PhD, Psychology, Royal College of Surgeons in Ireland, Dublin, UK

Previous research has focused on the relationship between negative emotions (e.g. depression) and QoL. However, less is known about the relationship between positive emotions (highlighted by the “positive psychology” movement) and QoL. For instance, is positive psychological well-being synonymous with QoL? Are subjective and psychological well-being similarly related to QoL? Well-being research consists of two broad traditions: psychological well-being (PWB) deals with human potential and subjective well-being (SWB) deals with happiness. Keyes (2002) provides a theoretical model of the relationship between these components. The aim of the study was to examine the relationship of PWB and SWB respectively with individual QoL (IQoL). IQoL was assessed using the Schedule for the Evaluation of Individual Quality of Life (SEIQoL). PWB (autonomy, environmental mastery, personal growth, positive relations with others, purpose in life, self-acceptance) was assessed using the Psychological Well-being Scale and SWB by the Satisfaction with Life Scale and the Positive and Negative Affectivity Scale. 136 university students completed the questionnaires. The mean SEIQoL index score was 65.05 (± 13.22 SD) (range 28-97; possible range: 0-100). Expected correlations were found between the SEIQoL index score and the PWB ($r=.26$ to $.43$, all $p<.01$) and SWB scales ($r=-.40$ to $.54$, all $p<.01$). Using structural equation modeling, Keyes' model was tested ($\text{CHI}^2=48.34$, $df=24$, $GFI=.92$, $CFI=.93$, $RMSEA=.08$) showing an acceptable fit. In a second step, the SEIQoL index score was entered into the model ($\text{CHI}^2=61.8$, $df=31$, $GFI=.92$, $CFI=.92$, $RMSEA=.08$). The model explained 41% of variance of the SEIQoL index (SWB $\beta=.83$, $p<0.001$; PWB $\beta=-.24$, $p=.44$). Correlations among measures suggest no redundancy of measures, nor that individual QoL is equivalent to either PWB or SWB. Results indicate that IQoL measured by SEIQoL primarily assesses SWB rather than PWB. These findings raise the question of whether, and to what extent, IQoL is primarily driven by SWB.

Session 7: Experience Sampling and Daily Process Analysis

EXPLORING PATIENT OUTCOMES AS DAILY PROCESSES: THE PROMISE AND THE CHALLENGE

Howard Tennen, PhD*, Professor, Department of Community Medicine and Health Care, University of CT Health Center, Farmington, CT

Patient outcomes emerge over time during the course of their everyday lives. Daily process outcomes research captures these temporally unfolding dynamics by emphasizing proximal events and experiences; measuring real-time change in rapidly fluctuating processes; minimizing random and systematic recall error; revealing individual differences in the temporal patterning of emotions, coping and symptoms; and strengthening our causal inferences. Through examples of daily process studies of chronic pain, alcohol use, and depression, Dr. Tennen will demonstrate how the application of daily process models enhance our description of outcomes, generate and test outcome-related hypotheses, and provide the basis for understanding the mechanisms of action of behavioral and pharmacological interventions.

DAILY MEASUREMENT DETECTS AN EARLIER ONSET ANTIDEPRESSANT EFFECT

William R. Lenderking, PhD, Outcomes Research, Mingxiu Hu, PhD, Biostatistics, Pfizer Inc, Groton, CT, Howard Tennen, PhD, Psychiatry, UCHC, Farmington, CT, Joe C. Cappelleri, PhD, Charles D. Petrie, PhD, Outcomes Research, Pfizer Inc, Groton, CT

The purpose of the study was to determine if daily assessment of depressive symptoms using date-verified, standardized, self-report measures completed at home would detect antidepressant effects earlier than weekly clinic-based assessments. Patients with MDD ($n=78$) received fluoxetine, were randomized to two assessment groups and followed for 28 days. The standard arm (SA) completed a battery of self-report and psychiatric assessments (e.g., HAM-D) at Baseline, Days 7, 14, and 28. The daily arm (DA) did the same weekly evaluations as the standard arm, plus a daily battery. We focus on three measures that were completed by both SA and DA: the Inventory of Depressive Symptoms-SR (IDS), the Schwartz Outcomes Scale (SOS-10), and a global rating (GRSD). Two binary outcomes were tested with survival analysis in responders: onset (20% improvement from baseline), and response

(50% improvement from baseline maintained at 35% through 28 days). HAM-D and MADRS scores were the same in both groups, indicating similar levels of depression and response. Among responders, using the IDS and the SOS-10, both onset and response were detected sooner in the DA (logrank $p < .007$ for all tests). Median times to onset and response were IDS onset: 4 vs 7 days; IDS response: 8 vs 28 days; SOS onset: 2 vs 7 days; SOS response: 4 vs 14 days. On the GRSD, time to onset was sooner in the DA ($p < .0001$), but response was not. Using the IDS and the SOS-10, we were able to detect both onset and response to antidepressant treatment among responders more rapidly using daily measures than using weekly clinic-based assessments alone. Since levels of depression were the same in both arms using the HAMD and MADRS, we believe our findings do not reflect measurement reactivity. Daily standardized assessments appear to be a sensitive method for detecting early response to antidepressant treatment, suggesting their potential application in future placebo-controlled depression studies as a reliable and valid measure of treatment effect.

DOES BETTER DATA FROM ELECTRONIC PATIENT REPORTED OUTCOMES (EPRO) METHODOLOGY ACTUALLY IMPROVE CLINICAL RESEARCH? RESULTS FROM A RANDOMIZED TRIAL COMPARING PAPER AND EPRO DIARIES

Stephen A. Raymond, PhD, Science & Research, PHT Corporation, Charlestown, MA, Jay D. Pearson, PhD, Epidemiology, Merck Research Laboratories, Blue Bell, PA

Use of electronic devices for patient self-reporting in clinical trials is increasing. Previous studies have shown ePRO to result in higher yields of data (higher percentage of completed, legible, attributable, logical and analyzable fields) and improved validity of time and date of data entry. However, it has not been established that these improvements actually lead to more conclusive results or more sensitive outcomes. In this study we tested the precision and power of paper vs. ePRO methods to measure 4 variables sleep latency (min), duration (min), quality and ability to concentrate (4 level Likert) in insomnia patients. Patients with chronic primary insomnia were randomized to two arms, not according to medication but according to the methodology used to capture data. The two arms, electronic vs paper, were stratified for age and education level. Patients in the paper arm (n=32) completed morning and evening self reports every day for a week using forms that were collected at each weekly visit. Patients in the electronic arm used handheld Palm® devices to capture the same data (PHT Corporation, Charlestown, MA) and transmit it to a central server every day. All patients underwent a 1 week washout period followed by 4 weeks of treatment for insomnia. Baseline values were each averaged for each patient during the washout week (baseline) and compared to the corresponding 7 day average in week 4 (treatment). Mean scores and treatment difference scores were similar for both arms. The key finding was that for all 4 variables, the variation was much LOWER for electronic capture (Latency by 20%, Duration by 36%, Quality by 25%, and ability to concentrate by 33%). The results provide an empirical example that demonstrates that higher integrity of data with ePRO methods yields greater statistical power (i.e. conclusive trials with 30% fewer patients) than paper methods do.

FEASIBILITY OF USING HANDHELD COMPUTERS FOR REAL-TIME ASSESSMENT OF FATIGUE IN CANCER PATIENTS

Karen M. Basen-Engquist, PhD, Behavioral Science, The University of Texas M. D. Anderson Cancer Center, Houston, TX, Carl de Moor, PhD, Biometry, The University of Texas School of Public Health, Houston, TX, Charles Levenback, MD, Judith Wolf, MD, Diane C. Bodurka, MD, Gynecologic Oncology, Cindy L. Carmack Taylor, PhD, Behavioral Science, The University of Texas M. D. Anderson Cancer Center, Houston, TX

Symptoms and quality of life are often measured retrospectively, which can lead to inaccuracies in the data caused by distorted recall, bias introduced by artificial testing situations or current emotions, and a lack of information related to patterns occurring over short time periods. This presentation will describe a pilot project to test the feasibility of using handheld computers for real time assessment of fatigue in cancer patients. This study recruited advanced ovarian cancer patients receiving carboplatin, paclitaxel, or both. Patients were asked to use the computers starting approximately one week before their next chemotherapy infusion and to record data throughout the next chemotherapy cycle. The computer prompted them to record information about their fatigue and other symptoms at four random times during the day, and 2 scheduled times (bedtime and wake time). 42 patients agreed to participate in the study. Of these, 4 became ineligible when their chemotherapy regimen changed. Five patients dropped out; 3 before starting to use the computer, and 2 because the computer was too disruptive. Of the 33 women who completed the study, 19 were newly diagnosed and 14 had persistent or recurrent disease. Their average age was 58.5 (range 27-18). Data completion rates were very high (86%); 20 of the 33 women completed over 90% of the assessments. To describe between-patient differences in fatigue patterns, fatigue was regressed on number of days since chemotherapy (for days 0-21) with separate regressions for each patient. A cluster analysis of each patient's

intercept (representing overall fatigue level) and slope (indicating change over time) yielded a 3-cluster solution suggesting three distinct patterns of fatigue: low fatigue throughout the chemotherapy cycle; fatigue throughout; and high fatigue at the beginning of the cycle followed by a dramatic decline. Using handheld computers for real-time collection of fatigue data is feasible and produced information on patterns of fatigue that could not be obtained with retrospective questionnaires.

EMOTIONS, STRESS AND HEALTH IN EVERYDAY LIFE

Alex Zautra, PhD*, Professor, Department of Psychology, Arizona State University, Tempe, AZ

In this talk, Professor Zautra expands on themes of resilience he introduced in his recent book, *Emotions, Stress, and Health* (Oxford University Press) through an examination of dynamic processes that influence the quality of everyday life. He will present data on how daily stressors, counterbalanced by positive interpersonal events influence emotional well-being and shape the capacity of patients so to sustain quality of life in the face of chronic pain.

Session 8: State of the Art Utilities/DCE/Q-TWiST

COMPARING AND CONTRASTING UTILITIES AND WILLINGNESS TO PAY

David Feeny, PhD*, Professor, Institute of Health Economics and University of Alberta, Edmonton, AB, Canada

The utility approach to assessing health-related quality of life (HRQL) and willingness to pay (WTP) are two widely used families of techniques for assessing preferences for health outcomes in the evaluation of health and healthcare interventions. Both approaches rely on stated preferences rather than revealed preference (what people actually do.) The conventional scale for utility scores assigns a value of 1.0 to perfect health and 0.0 to dead. Utility scores obtained by direct assessments using techniques such as the standard gamble or by using multi-attribute systems are used to weight the time spent in each health state to estimate quality-adjusted life-years gained (QALYs). Estimates of the QALYs gained are in turn used as the denominator in cost-utility analyses (CUA). Results are reported as incremental net cost per QALY gained. Similarly WTP estimates are used to assess the value of the gain expressed in pecuniary (dollar) values. These estimates support cost-benefit analyses (CBA) in which all of the costs and benefits are expressed in dollar values and the results are reported as net benefits. The presentation will compare and contrast these two approaches with respect to a number of characteristics including the scope and comprehensiveness of the evaluation, nature of the objective function, informational and cognitive burdens on respondents, and ability to capture the underlying structure of preference judgments. Conceptual and practical advantages and disadvantages of both approaches will be discussed. The focus will be on “best practice” for assessing HRQL effects for both approaches. Conceptual advantages of WTP may be attenuated by practical disadvantages.

BRIDGES AND BARRIERS: WILLINGNESS TO PAY, WILLINGNESS TO WAIT, AND HEALTH-STATE UTILITY MEASURES

F. Reed Johnson, PhD*, Principal Economist and Senior Fellow, RTI Health Solutions, Research Triangle Institute, Research Triangle Park, NC

Health economists and other economists measure utility in different ways. While QALYs are a powerful and useful way to compare dissimilar health outcomes, conventional economists often are puzzled by the weak link between such measures and the individual welfare measures used in nearly every other area of applied economics. This presentation discusses opportunities for bridging this gap and possible barriers that may stand in the way of progress.

Q-TWiST: DO'S AND DON'TS

Richard D. Gelber, PhD, Prof, Dept. of Biostatistical Science, Dana-Farber Cancer Institute, Boston, MA, Benard F. Cole, PhD, Assistant Professor of Community and Family Medicine, Dartmouth University, Hanover, NH

Q-TWiST (Quality-adjusted Time Without Symptoms of Disease and Toxicity of Treatment) is a statistical method designed to highlight treatment effect tradeoffs between quantity and quality of life. Data obtained directly from clinical trials are used. Three steps are required to perform a Q-TWiST analysis: (1) Define quality-of-life oriented clinical health states that reflect the disease and treatment specific trade-offs being evaluated; (2) Estimate the average amount of time patients spend in each of the clinical health states using Kaplan-Meier methods; (3) Compare the treatments using a threshold utility plot to determine the range of utility coefficient values (assigned to weight the different clinical health states) for which one treatment is preferred to the other. The Q-TWiST statistic is the weighted average of the clinical health state durations. Most often, some observations in the clinical trial are

**invited presenter*

censored at the time of analysis. In this case, biased estimates of average Q-TWiST are obtained if the Q-TWiST statistics for individual patients are used in a Kaplan-Meier calculation. Unbiased estimates of clinical health state durations in step (2) must be obtained from the areas between Kaplan-Meier curves calculated for nested transition times from one health state to the next. Examples and further perspectives will be presented.

ATRASENTAN INCREASES QUALITY ADJUSTED TIME TO PROGRESSION IN MEN WITH HORMONE-REFRACTORY PROSTATE CANCER METASTASIZED TO BONE

Parvez M. Mulani, PhD, Talat Ashraf, MD, Center for Pharmaceutical Appraisal and Outcomes Res, Andrew Allen, MD, PhD, Oncology Franchise, Sandra Hulting, MS, Jeffrey Isaacson, PhD, Clinical Statistics, Darryl Sleep, MD, Oncology Franchise, Abbott Laboratories, Abbott Park, IL

Introduction: Quality of Life (QoL) is an important factor in the evaluation of new cancer therapy. The objective of this study was to assess the benefit of atrasentan on disease progression after balancing the undesirable side effects and the impact on QoL. This was achieved by integrating temporal benefit and QoL into a single analysis of patients with hormone refractory prostate cancer (HRPC) metastasized only to bone from a phase 3 study of atrasentan. **Methods:** The European Organization for Research on the Treatment of Cancer Quality of Life Questionnaire (EORTC QLQ C-30) and the Functional Assessment of Cancer Therapy (FACT) were completed at baseline, week 4, week 12, and then every 12 weeks. QoL scores were converted to a unit scale; (0-1), with a higher score indicating better QoL. The quality adjusted time to progression (QATTP) was computed as the sum of the product QoL weights and the duration for which the patient experienced that level of QoL. The QATTP was analyzed by Cox proportional hazards methodology. **Results:** Of 473 men presenting with metastases only to bone, 220 and 253 received placebo and atrasentan 10 mg QD, respectively. The QATTP was significantly longer for atrasentan than for placebo for all of the QoL domains analyzed, and the treatment difference was greater than the unadjusted time to progression (TTP) (Table:QATTP Results). **Conclusions:** The benefit of delayed disease progression observed with atrasentan is greater when TTP is estimated jointly with QoL. Atrasentan results in a significantly delayed QATTP, and the treatment difference is greater than that seen in the unadjusted TTP analysis, indicating that patients with HRPC metastases confined to bone also derive a QoL benefit from treatment with atrasentan.

QATTP Results			
	Hazard Ratio	95% CI	Log Rank P Value
Unadjusted TTP	1.40	1.14 - 1.72	0.002
QOL Domain			
FACT			
FACT-P	1.41	1.15 - 1.74	0.001
Prostate Specific Scale	1.41	1.14 - 1.74	0.001
Composite Score	1.42	1.15 - 1.75	0.001
EORTC			
Pain	1.45	1.17 - 1.79	<0.001
Global Score	1.46	1.18 - 1.80	<0.001

TIME TRADE-OFF VALUATIONS OF EQ-5D HEALTH STATES: ARE THE US AND UK DIFFERENT?

Jeffrey A. Johnson, PhD, Nan Luo, PhD, Public Health Sciences, University of Alberta, Edmonton, AB, Canada, James W. Shaw, MS, Pharmaceutical Sciences, University of Arizona, Tucson, AZ, Paul Kind, MPhil, Centre for Health Economics-Outcomes Research Group, University of York, Heslington, York, UK, Stephen J. Coons, PhD, Pharmaceutical Sciences, University of Arizona, Tucson, AZ

Purpose: The purpose of this study was to compare directly elicited time trade-off (TTO) valuations for EQ-5D health states between the US and UK general adult populations. **Methods:** We analyzed data collected in the US and UK EQ-5D valuation studies. Using the same TTO protocol, valuations of 42 common EQ-5D health states were elicited from representative samples of the US and UK general adult populations. We estimated mean TTO valuation scores for each health state for both populations and compared these mean scores between the two countries. We then conducted random effects regression modeling to examine country-specific difference in the valuations while adjusting for effects of other known predictors (i.e., age and sex). **Results:** Estimated mean valuation scores of the 42 health states ranged from -0.38 to 0.88 for the US population and from -0.54 to 0.88 for the UK population, with the US mean scores being higher than UK mean scores for 41 health states (mean difference: 0.11; range of difference: 0 to 0.25). In regression analysis, we confirmed the country-specific differences in valuations (coefficient 0.09; $p < 0.001$) and found that magnitude of the differences was associated with characteristics of health states valued. For example, compared with its UK counterpart, the US population rated health states characterized by extreme pain/discomfort 0.08 points higher, but health states characterized by moderate problems in mobility were valued 0.03 points higher compared to the UK. **Conclusions:** Meaningful differences exist in directly elicited TTO valuations of EQ-5D health states between the US and UK general adult populations. The differences in valuations depend on the degree of severity of the health states and the specific dimensions affected.

Abstracts for Posters

Abstract 100

PROOF OF CONCEPT OF A WIRELESS APPROACH FOR ENABLING COMMUNICATIONS BETWEEN GERMAN PHYSICIANS AND THEIR PATIENTS WITH TYPE 2 DIABETES TREATED WITH NATEGLINIDE

Paul M. LeVine, MS, Analytic Services, David R. Netherton, MS, MA, Senior Director, Global Programs, InfoMedics, Inc., Woburn, MA

Purpose of Research: We report on the results of proof-of-concept study designed to allow German physicians treating patients with Type 2 diabetes with nateglinide to gain patient feedback using an entirely wireless design. **Subject Sample:** 60 patients and 5 physicians agreed to participate in the study. **Statement of Methods:** Using mobile communications devices, patients reported on their medication-taking experience and clinical values, including blood glucose levels, according to a clinical protocol have five (5) patient reporting events during the study period of 14 days. This information was then delivered via dedicated web technology to treating physicians. **Summary of Results:** Participating physicians enrolled an average of seven patients in the study. Each physician visited the website an average of 40 times. 21% of these visits involved the physician reviewing individual patient responses. 60% of physicians agreed that their understanding of the medication and of the condition was improved by their participation in the study. 65% of patients reported believing that the wireless handset device was easy to use, and 71% noted that the screen of the device was acceptable for reading, navigating and entering information about their condition. **Conclusions:** In a small proof-of-concept study designed to assess the viability of entirely automated communications among physicians and patients regarding treatment and clinical endpoints in Type 2 diabetes, both physicians and patients were able to use the system without significant difficulty and reported favorable experiences with the approach.

Abstract 101

USING COMPUTERIZED ADAPTIVE TESTING IN HEMODIALYSIS PATIENTS

Asha P. Masurekar, BA, Outcomes Monitoring Program, Dialysis Clinic, Inc., Boston, MA, Michelle M. Chapman, PharmD, Nephrology, Tufts-New England Medical Center, Boston, MA, Anastasia Raczek, PhD Candidate, Measurement Science, QualityMetric, Inc., Lincoln, RI, Klemens B. Meyer, MD, Nephrology, Tufts-New England Medical Center, Boston, MA, John E. Ware, Jr., PhD, Measurement Science, QualityMetric, Inc., Lincoln, RI

Kidney failure significantly affects patients' quality of life, comorbidities are common, mortality is high and treatment options are limited. Routine health status assessment in dialysis patients has used the fixed version of the SF-36. However, fixed surveys are burdensome, costly, and have limiting ceiling and floor effects. Computerized Adaptive Testing (CAT) may address these problems. We field tested CAT among hemodialysis outpatients, evaluating the acceptance and feasibility. Using confirmatory factor analysis and Item-Response Theory (IRT), 3 scales, Sleep (7 items), Pain (39 items) and The Effects of Kidney Disease (17 items) were developed. Real-data CAT simulation analyses, using actual data from previous studies of patients with Chronic Kidney Disease (CKD), were completed in order to compare multiple adaptive scale versions with the IRT latent trait score estimated by the full model and to understand the merits of alternative strategies for programming the CKD-CAT. Eligible hemodialysis patients were adults without dementia, in 2 centers in Boston, who spoke English, and had good vision. A fixed survey (63 items) and CAT were completed in random order. CAT used a laptop computer and a hand held mouse. Of 91 patients approached, 49 (age 57.8 ± 15.1 years, 38% women, 69% white) participated. Half of the participants had minimal or no previous computer experience. CAT precision standards were met at mean 15.2 ± 1.5 items; the fixed survey had 63 items. The CAT also took less time (6.6 ± 4.9 min) than the fixed survey (14 ± 6.6 min). Patients that became frustrated with the CAT (32%) had study staff mark their responses. More patients preferred CAT (57%); 25% preferred the fixed survey. In this pilot study, we found CAT to be a feasible alternative to fixed surveys in hemodialysis patients. CAT reduces respondent burden and the administration time as compared to fixed surveys. Despite lack of computer experience, a majority of patients preferred CAT, demonstrating its clinical acceptability.

Abstract 102

LESSONS LEARNED FROM COMPUTERIZED ADAPTIVE TESTING ADMINISTRATION IN HEMODIALYSIS PATIENTS

Asha P. Masurekar, BA, Outcomes Monitoring Program, Dialysis Clinic, Inc., Boston, MA, Michelle M. Chapman, PharmD, Nephrology, Tufts-New England Medical Center, Boston, MA, Anastasia Raczek, PhD Candidate, John E. Ware, Jr., PhD, Measurement Science, QualityMetric, Inc., Lincoln, RI, Klemens B. Meyer, MD, Nephrology, Tufts-New England Medical Center, Boston, MA

In 2 centers in Boston, English speaking adult hemodialysis patients without dementia, and with vision good enough to read a computer screen, completed a fixed survey and a Computerized Adaptive Test (CAT) in random order. The CAT used a laptop computer and a hand held mouse. Patients undergoing hemodialysis are tethered to a machine and are limited in movement for hours. No published information exists on computer use in this population. We report qualitative observations and the results of a semi-structured survey. Of 91 patients approached, 49 (age 57.8±15.1 years, 38% women, 69% white) completed both surveys. CAT precision standards were met at 15.2±1.5 items; the fixed survey had 63 items. CAT also took less time (6.6±4.9 min vs. 14±6.6 min). Study staff observed that dialysis treatment hindered computer use. The treatment chair design limited space to manipulate a mouse, which also became entangled with some patients' dialysis tubing. Half of the participants had little or no prior computer experience. Gripping the mouse for an extended time resulted in painful hand cramps. Patients that became frustrated with the CAT (32%) asked study staff to mark their responses. Patients instinctively touched the computer screen to indicate their answer. Changes in blood pressure automatically detected by dialysis machines set off alarms, interrupting survey administration. We enlarged the screen resolution to display approximately 14-point font for 42% of patients because they needed larger sized type than the 10-point font of the CAT. Although CAT reduced patient burden, this pilot project uncovered unexpected administration issues. To assess the efficacy of CAT in chronic diseases, technology must meet the needs and characteristics of the specific population. Pilot tests in different populations should be conducted to ensure large-scale success.

Abstract 103

DIFFERENTIAL ITEM FUNCTIONING: USEFULNESS OF DIF IN QOL RESEARCH

Linda G. Marc, MPH, MS, Lisa F. Berkman, PhD, Society, Human Development and Health, Harvard School of Public Health, Boston, MA, Charles Lewis, PhD, Psychology, Fordham University, Bronx, NY, Marcia A. Testa, PhD, Biostatistics, Harvard School of Public Health, Boston, MA

Most quality-of-life (QoL) measures are developed and validated in well-educated and non-minority populations. Consequently, there is potential for measurement bias across population groups, limiting the generalizability of QoL measures. Although DIF is used extensively in the area of educational measurement, it is underutilized in QoL research. Using the HIV MOS Scale, DIF methodology is applied to identify items that behave unexpectedly relative to overall group performance. Groups are defined by educational attainment (EA) in this study population. Baseline data are from the Adult AIDS Clinical Trial Group, Protocol-384 (ACTG-384), which enrolled randomized antiretroviral-naïve individuals (N=980) infected with HIV-1 to three- and four-drug regimens between 1998-99. The predictor, EA was measured as an ordinal categorical variable (<HS =15%, HS=30%, Some College/2yrs = 31%, 4yrs-College/Graduate Degree = 24%). Subjects were predominately male (81.5%), 46.5% Caucasian, and 35.0% Black and with mean age 37 years. The Zumbo-Thomas Method for DIF is illustrated, using ordinal logistic regression (OLR) to predict item responses. Modeling steps consist of: (1) controlling for respondents from different groups according to their total item scores; (2) assessing the presence of uniform DIF "item/group preference" (confounding); and (3) non-uniform DIF "performance/group interaction" (effect modification). Model-building strategies include the likelihood ratio and chi-square difference tests within nested OLR models; tests of significance (p-values) and comparing measures of effect (McKelvey-Zavoina R-square and Odds Ratio). Group differences in item performance do not necessarily constitute item bias, as responses may be valid reflections of the knowledge/experiences between groups. DIF methods contribute to QoL research by drawing attention to differential item endorsement. This may adversely influence the course of clinical treatments and policy decisions across population groups.

Abstract 104

A NEW TOOL FOR MONITORING PEDIATRIC HEALTH OUTCOMES: THE SF-10 FOR CHILDREN™

Diane M. Turner-Bowker, PhD, Martha S. Bayliss, MSc, Research & Evaluation, QualityMetric, Inc., Lincoln, RI

Pediatric health-related quality of life (HRQOL) measures that are comprehensive, reliable, valid, easy to score and interpret, and minimize respondent burden are needed to quantify the impact of health on functioning. This research focused on the development of the SF-10 for Children™, a brief 10-item parent-completed tool designed to measure physical and psychosocial functioning of children ages 5 and older. The SF-10 for Children is an alternate short-form to the Child Health Questionnaire™ (CHQ), developed in the early 1990s as a generic measure of pediatric HRQOL. Development of the SF-10 for Children used cross-sectional CHQ data (N=411) from the National Survey of Functional Health Status. Items from the pool (k=50) were categorized by the CHQ content domain and submitted to a series of forward stepwise regressions. Results were evaluated by how consistently items (independent variables) entered the model as strong predictors of the criterion dependent variables (summary scale scores). Ten items emerged as strong predictors of physical (PHS) and psychosocial (PSS) summary scales. Collectively, these 10 items explain 88% of the variance in the CHQ PHS score, and 89% of the variance in the CHQ PSS measure. Item-level tests of discriminant and relative validity were used to select the shortest set of items to represent the new short-form survey. Scale-level discriminant and relative validity tests were completed using the same criteria for item-level analyses. Results show that the SF-10 for Children retains the validity of the CHQ with substantially less respondent burden. Most items correlated 0.40 or higher with the hypothesized scale, and passed discriminant validity tests. Internal consistency reliability met the minimum of 0.70 for group comparisons (Cronbach alpha=.70, .72 for PHS and PSS respectively). There were no floor effects, but moderate ceiling effects for PHS, not unexpected for a general population sample. Development of the SF-10 for Children was a data-driven process resulting in an empirically sound, useful tool for measuring impact on HRQOL in a pediatric population.

Abstract 105

MORE SENSITIVE HAMD SUBSCALES FOR ANTIDEPRESSANT TRIALS

Mingxiu Hu, PhD, Shuang Wang, PhD, Development Operations, Pfizer Inc, New London, CT, Michael Krams, MD, Clinical Research, Pfizer Inc, Groton, CT

Hamilton Depression Rating Scale (HAMD) has served as the primary standard to quantify depression severity in antidepressant clinical trials. Many experts believe that HAMD is not a sensitive instrument, which is partially responsible for the high failure rate of antidepressant trials. Myriad efforts have been made to develop more sensitive HAMD subscales for detecting treatment effects. Bech's six-item, Maier's six-item, and Gibbon's eight-item scales are the products of these efforts. Faries et al (2000) and Entsuah et al (2001) compared the sensitivity of the HAMD total and the three subscales using a number of studies and concluded that the three unidimensional subscales are more sensitive than the HAMD-17 total for detecting treatment effects. This presentation describes an even more sensitive subscale (with nine HAMD items) that we established through an exclusive searching algorithm by evaluating the effect sizes and t-testing statistics for all possible combinations of 4 to 9 items. Four studies were used to build the subscale and one is for cross-validation. The study sizes (two treatment arms combined) vary from 117 patients to 191 patients. The proposed subscale, HAMD total, and the other three subscales have comparable Cronbach's alpha, which is commonly used to measure internal consistency of a questionnaire. On the other hand, this subscale was at least as sensitive (in terms of t-testing statistics) as the HAMD total and the three well-established subscales across all analyses: The four original studies, cross-validation, and a meta-analysis. An Item Response Theory (IRT) analysis using software MULTILOG also showed acceptable results for the subscale.

Abstract 106

LINKING SELF-ESTEEM, CONFIDENCE, AND RELATIONSHIPS WITH SEVERITY OF ERECTILE DYSFUNCTION

Joseph C. Cappelleri, PhD, Clinical Research, Pfizer Global Research & Development, Groton, CT, Sandeep Duttgupta, PhD, Global Outcomes Research, Nancy Sherman, Therapeutic Operations, Vera J. Stecher, PhD, Richard L. Siegel, MD, Li-Jung Tseng, PhD, US Pharmaceuticals Group, Pfizer Inc, New York, NY

We assessed changes in the psychosocial factors of self-esteem, confidence, and relationships after treatment with placebo or sildenafil citrate in men with erectile dysfunction (ED). This was an international, multi-center, 12-week, double-blind, flexible-dose (25, 50, 100 mg) study involving 300 men (placebo, n=149; sildenafil, n=151) with ED from Brazil, Mexico, Australia, and Japan. Self-esteem, confidence, and relationship changes were measured from baseline to end of treatment (EOT) with the validated 14-item ED-specific Self-Esteem And Relationship (SEAR) questionnaire (score range: 0-100, higher scores more favorable). Severity of ED was based on EOT scores from the 6-item erectile function (EF) domain of the International Index of Erectile Function (score range, 1-30). Change scores on the SEAR were regressed on severity grades of the EF domain at EOT using an ANCOVA model adjusted for center and baseline SEAR score. Across treatment groups, higher levels of EF at EOT were linked generally to favorable mean changes in SEAR scores. Mean changes in all SEAR scores differed meaningfully and significantly (P<0.0001) across ED severity groups at EOT (Table). At EOT, more sildenafil-treated men had higher levels of EF compared with placebo-treated men (sildenafil vs placebo): severe ED, 8% (n=12) vs 32% (n=45); moderate ED, 4% (n=6) vs 13% (n=18); mild-moderate ED, 8% (n=12) vs 13% (n=18); mild ED, 18% (n=26) vs 8% (n=11); no ED, 61% (n=88) vs 34% (n=47). Improvements in self-esteem, confidence, and relationships were linked to higher levels of EF. The discriminating ability of the SEAR is further validated in making distinctions among grades of ED severity.

Abstract 107

SEAR Component	Overall Change	Change in Placebo	Change in Sildenafil	Correlation
1. Sexual Relationship	38.0	17.1	20.9	0.33*
2. Confidence	44.0	20.0	24.0	0.22*
3. Self Esteem	48.0	27.0	21.0	0.35*
4. Relationship	47.0	24.0	23.0	0.52**
Overall score	44.0	23.0	21.0	0.32*

SELF-ESTEEM, CONFIDENCE, AND RELATIONSHIPS IN MEN WITH ERECTILE DYSFUNCTION TREATED WITH SILDENAFIL OR PLACEBO: A COMBINED ANALYSIS OF 2 DOUBLE-BLIND, PLACEBO-CONTROLLED TRIALS
 Joseph C. Cappelleri, PhD, Clinical Research, Pfizer Global Research & Development, Groton, CT, Sandeep Duttagupta, PhD, Jessica Bowler, PhD, Arthur R. Crowley, MD, Richard L. Siegel, MD, Li-Jung Tseng, PhD, Global Outcomes Research, Pfizer Inc, New York, NY

We used the validated, erectile dysfunction (ED) specific Self-Esteem And Relationship (SEAR) questionnaire to assess changes in self-esteem, confidence, and relationships in men with ED treated with sildenafil or placebo. Two 12-week, double-blind, placebo-controlled, flexible-dose (25, 50, or 100 mg) trials were conducted in men aged >=18 y with clinically documented ED. Patients receiving nitrates were excluded. Change scores from baseline to end of treatment on the SEAR questionnaire were analyzed between treatment groups using linear regression. Pearson's correlation coefficients were obtained on changes in SEAR components with changes in the erectile function (EF) domain of the International Index of Erectile Function (IIEF). 274 (mean age 55 y; mean ED duration 4.3 y) and 279 patients (mean age 56 y; mean ED duration 4.4 y) received placebo or sildenafil, respectively. Sildenafil produced significantly greater improvements in all SEAR components (*P<0.0001) and all 5 domains of the IIEF. Changes in SEAR scores (possible range: -100 or least favorable to 100 or most favorable) showed moderate-to-high correlations with changes in EF domain scores (**P<0.0001). The most frequent adverse events (sildenafil vs. placebo) were headache (13% vs 6%), rhinitis (5% vs 1%), dyspepsia (5% vs 2%), and flushing (8% vs 1%). In men with ED, sildenafil produced substantial improvements in self-esteem, confidence, and relationships. Changes in these psychosocial factors correlated significantly with efficacy as determined by changes in the EF domain.

Abstract 108

A COMBINED EMPIRICAL / CONCEPTUAL APPROACH TO DETERMINING CLINICALLY MEANINGFUL SUBSCALES IN THE DEVELOPMENT OF A QUALITY OF LIFE MEASURE: A RECTAL CANCER EXAMPLE

Jennifer Bacik, MS, Howard T. Thaler, PhD, Epidemiology and Biostatistics, Douglas W. Wong, MD, Larissa Temple, MD, Surgery, Memorial Sloan-Kettering Cancer Center, New York, NY

Assessment of patient function and quality of life (QOL) has become an essential element of outcome evaluation in many clinical trials. Development of a psychometrically validated tool to measure patient function, however, is a difficult process due to the high dimensionality and subjective nature of QOL data. When presented with these data, the statistician's major challenge is to determine the underlying factor structure, including the appropriate number of factors, while ensuring that the resulting subscales are clinically meaningful. Factor analysis of variance (FANOVA) is a useful method when a global effect is present. The traditional "eyeballing" of the scree plot can be enhanced by a comparison of the eigenvalues obtained from the original data to the distribution of eigenvalues obtained from random permutations of the original data. Redundant or uninformative items need to be pruned based on correlations and clinical relevance. Additional confirmation of the appropriateness of the chosen subscales can be obtained by correlating the subscale scores with other QOL measures (construct validity), and by evaluating the subscales' ability to differentiate between clinically relevant groups of patients (discriminant validity) using a priori hypotheses established by the clinician. This combined empirical / conceptual approach is an appropriate paradigm for the determination of the underlying factor structure of the data. It is important that the statistician and clinician collaborate to ensure that the resulting tool is psychometrically sound and clinically meaningful. This approach is illustrated with data collected on patients undergoing sphincter-preserving surgery for rectal cancer.

Abstract 109

DEVELOPMENT OF A COMPUTERIZED ADAPTIVE TEST FOR STRESS (STRESS-CAT) AND ITS CLINICAL APPLICATION

Rueya D. Kocalevent, MA, Otto B. Walter, MD, Janine Becker, MA, Herbert Fliege, PhD, Burghard F. Klapp, MD, Psychosomatic Medicine, University Hospital Charité, Berlin, Germany, Jakob B. Bjorner, PhD, MD, QualityMetric, Inc., Lincoln, RI, Matthias Rose, MD, Psychosomatic Medicine, University Hospital Charité, Berlin, Germany

Psychological stress concepts emphasize the individual perception and appraisal of a situation as stressful. Measuring perceived stress has concentrated on two basic dimensions, environmental stressors and stress reactions. The study aims to develop and validate computerized adaptive tests (CAT) to measure these two stress dimensions in the clinical field. 1092 psychosomatic patients answered 140 items out of 4 validated stress questionnaires (PSQ, TICS, ABF, SUB) within their diagnostic routines at the University Clinic in Berlin between 2000-2002. 5 independent reviewers rated 47 out of the available 127 items as representative of the stressor dimension and 39 as representative of the stress reaction dimension. Exploratory factor analyses yielded two dimensions, stressor and stress reaction. The two dimensions were analyzed separately. All items underwent six steps of analyses to check for the assumptions of unidimensionality, local independence and item discrimination, using the programs Mplus, TestGraf, SAS, and Parscale. The Generalized Partial Credit Model could be applied, showing a good model fit. The CAT scores were estimated using an "expected a posteriori" algorithm (EAP). We could demonstrate the latent trait (stress) can be estimated with approximately 7 items +/- 2 items (mean +/- sd) on a highly precise level (reliability .90) with high correlation to established instruments based on classical test theory. Accordingly the absolute number of items could be substantially reduced (-31% to -77%) without loss of major information and with improved measurement precision in high and low score ranges. Measuring stress in the clinical field can be done at the same time economically and with high precision using a individually "tailored" computer adaptive test.

Abstract 110

USE OF PERSONAL DIGITAL ASSISTANTS (PDAS) FOR PSYCHOMETRIC ASSESSMENT AND COMPUTER ADAPTIVE TESTING

Otto B. Walter, MD, Matthias Rose, MD, Medical School of Humboldt University Berlin, Charite, Clinic for Psychosomatics, Berlin, Germany

Computer adaptive tests (CATs) exhibit a number of advantages in comparison to paper-and-pencil questionnaires. However, most current implementations of CATs are bound to PCs and therefore require rather expensive hardware and impose inconveniences on investigators and patients alike. We investigated applicability and costs of PDAs in psychometric diagnostics in comparison to traditional approaches to psychometric assessment. From 1990 to 1995, all patients of the Clinic for Psychosomatic Medicine at the Charité were investigated with a set of standardized paper-and-pencil questionnaires as part of our diagnostic routine. In 1995, we replaced paper-and-pencil questionnaires with a computer assisted assessment on PDAs, and in 2003 CATs for depression, anxiety, and stress (stressor and stress reaction) were integrated into our routine. A total of N=9659 patients were analyzed with respect to these three different approaches to psychometric assessment. Factor analysis of paper-and-pencil questionnaires and PDA-based assessment showed almost identical psychometric properties. Acceptance of PDA-based assessment was high among patients. Similar to the paper-and-pencil version, less than 1% of the patients rejected the examination. The use of PDAs reduced costs of our psychometric assessment by two thirds. Apart from these economical advantages, another beneficial effect of PDAs was the possibility to collect data for the construction of CATs. These tests were developed from responses from 3270 (Depression-CAT), 2348 (Anxiety-CAT), and 1432 (Stress CATs) patients and could be implemented on the PDA platform. Even though a further cost reduction could not be observed, the introduction of CATs reduced the burden placed on patients as constructs can now be measured with markedly fewer items. For instance, on average 7 items are sufficient to measure anxiety with high precision ($SE < .32$) with our Anxiety-CAT. This is one third of the items required by conventional testing. Our results indicate that both clinical diagnostics and research designs can benefit from inexpensive and mobile data collection with PDAs.

Abstract 111

VALIDATION STUDY OF THE GERMAN COMPUTERIZED ADAPTIVE TEST FOR ANXIETY ('ANXIETY-CAT')

Janine Becker, PhD, Otto B. Walter, PhD, Herbert Fliege, PhD, Ruya D. Kocalevent, Dipl.-Psych., Psychosomatic Medicine, Medical Clinic, Charité, Berlin, Germany, Jakob B. Bjorner, PhD, Deputy Chief Science Officer, QualityMetric, Inc., Lincoln, RI, Burghard F. Klapp, Prof, Matthias Rose, PhD, Psychosomatic Medicine, Medical Clinic, Charité, Berlin, Germany

Recently a German Computerized Adaptive Test has been developed on principles of the Item Response Theory (IRT) to measure for anxiety ('Anxiety-CAT') in healthy persons, patients with chronic somatic diseases, as well as psychosomatic and psychiatric in- and outpatients. First simulation studies reanalyzing computer assisted response data of $n = 2.348$ patients demonstrate the underlying item bank (50 items) to offer the potential of a very precise ($SE \geq 0.32$) and efficient measurement of anxiety (average number of items used: $7+/-2$ (mean+/-sd). The purpose of the current study was to examine if the results can be replicated in a prospective design and to investigate the validity of the developed instrument. For evaluating the validity, the 'Anxiety-CAT' and several established psychometric instruments measuring anxiety (HADS, BAI), and personality (NEO-PI, GT) as well as a computer-assisted standardized diagnostic interview (CIDI) were administered to a sample of $n = 102$ psychosomatic inpatients being treated between 10/02 and 10/03. The results of the prospective design are in line with the results of the simulation study in that only $5.3+/-1.9$ (mean+/-sd) were needed in order to measure anxiety on a highly precise level ($SE \geq 0.32$). CAT scores correlated well with anxiety scores of conventional instruments (HADS 'Anxiety Scale': $r=0.77^{**}$; BAI $r=0.55^{**}$; NEO-PI-'Neuroticism' Scale: $r=0.63^{**}$). Furthermore the results point at a good discriminant validity to different scales of personality (range: $r=-0.07$ to $r=-0.21$). Validating the CAT scores by the diagnoses as administered by the standardized diagnostic interview (CIDI) revealed significantly higher CAT scores for inpatients with anxiety disorders than for inpatients without mental disorders or for a small sample of students ($n=35$). But discriminating between different diagnoses of mental disorders by CAT scores was quite difficult due to large comorbidity. Overall this validation study revealed promising psychometric properties, but further research on larger non-clinician samples is needed.

Abstract 112

QUALITY OF LIFE (QOL) RESPONSE: GRADING AND MEASUREMENT

Andrei A. Novik, Prof, National Pirogov Medical Surgical Center, National Pirogov Medical Surgical Center, Moscow, Russia, Tatiana I. Ionova, PhD, Multinational Center of QOLR, Multinational Center of QOLR, St. Petersburg, Russia, Konstantin V. Liadov, Prof, National Pirogov Medical Surgical Center, National Pirogov Medical Surgical Center, Moscow, Russia, Anton V. Kishtovich, PhD, Anna A. Tzepakova, PhD, Multinational Center of QOLR, Multinational Center of QOLR, St. Petersburg, Russia

The final goal of treatment of chronic diseases is to improve a patient(s) QoL. Evaluation of QoL response along with a clinical response might be strongly recommended for patient management. According to our approach there are 4 grades of QoL response: complete QoL response (CR); partial QoL response (PR); QoL stabilization; QoL worsening. The aim of the study is to develop the algorithm of measuring QoL response in patients with rheumatoid arthritis (RA). 205 RA patients (males/females-29/176, mean age 54.3) in acute disease phase were enrolled in the study. Patients filled in SF-36 at baseline and after 6 weeks of treatment. The integral QoL index (IQLI) was calculated using the method of integral profiles (Novik, Ionova et.al.,2003). To compare QoL parameters of RA patients with population norms a sample of 428 respondents adjusted by age and gender was used. Patients and norm populations were stratified based on IQLI. Comparison between groups was conducted by Mann-Whitney U-test or Wilcoxon test. Stratification of normal population revealed five groups with IQLI means of 0.10(n=74); 0.26(n=159); 0.44 (n=159);0.44(n=128);0.64(n=46);0.81(n=21), respectively. Five groups of patients with RA at baseline were also observed with IQLI means of 0.001(n=7);0.06(n=126);0.16(n=42);0.29(n=19);0.49(n=11), respectively. After 6 weeks of treatment, IQLI in these groups appeared to be 0.05;0.18;0.27;0.31;0.53. Comparison of IQLI in each patient group before and after treatment revealed significantly important QoL improvement. Comparison of IQLI between corresponding groups of controls and patients after treatment showed a statistically significant difference. PR was registered, no CR, QoL stabilization and QoL worsening took place. Because of heterogeneity of both patients and norms, we propose to focus the analysis on differences in mean values of IQLI between corresponding groups of treated and normal individuals. Four grades of QoL response are proposed.

Abstract 113

PARADIGM AND ALGORITHM OF CLINICAL SIGNIFICANCE: EVALUATION OF HEALTH RELATED QOL DATA

Andrei A. Novik, Prof, National Pirogov Medical Surgical Center, National Pirogov Medical Surgical Center, Moscow, Russia, Tatiana I. Ionova, PhD, Multinational Center of QOLR, Multinational Center of QOLR, St.Petersburg, Russia, Konstantin V. Liadov, Prof, National Pirogov Medical Surgical Center, National Pirogov Medical Surgical Center, Moscow, Russia, Gary I. Gorodokin, MD, Multinational Center of QOLR, Multinational Center of QOLR, Brooklyn, NJ

Interpretation of HrQOL scores depends on a number of clinical factors such as disease type, stage of disease, patient sensitivity etc. It is really difficult to analyze all of them and to propose a universal decision making model for practicing medicine based on clinical and HrQOL data. We have analyzed HrQOL data of patients with various types of disorders (about 320 patients with multiple sclerosis, rheumatoid arthritis, systemic lupus erythematosus and lymphomas) at different stages of treatment and at follow-up and developed a paradigm and an algorithm of clinical significance evaluation of HrQOL data which might be applied to everyday clinical practice. The paradigm is based on three major parts: - Clinical relevance of HrQOL data. - Clinical importance of HrQOL data. - Clinical interpretation of HrQOL data. According to the algorithm the physicians should perform three logic steps following the paradigm parts: 1. To analyze the clinical relevance of HrQOL data regarding disease type and stage, as well as the treatment approach. 2. To define the clinical importance of HrQOL data taking into consideration that the absence of changes in HrQOL scores during treatment might be really important for many patients, suffering from chronic diseases. 3. To give a clinical interpretation of HrQOL data. Comprehensive analysis of biological and physiological parameters along with HrQOL data allows to explain the mechanisms of formation of disturbed HRQOL patient(s) profile. The proposed paradigm and an algorithm of clinical significance evaluation of HrQOL data allow clinicians to create a critical background for decision making in practicing medicine based on important internal medicine principle - to treat not only the disease, but the patient as well.

Abstract 114

THE RELIABILITY AND VALIDITY OF THE CZECH VERSION OF A BRIEF MENTAL HEALTH OUTCOME MEASURE

Eva Dragomirecka, PhD., Social Psychiatry, Prague Psychiatric Center, Prague, Czech Republic, William R. Lenderking, PhD, Outcomes Research, Pfizer Inc, Groton, CT, Lucie Motlova, MD, School of Public Health, University of California, Berkeley, CA, Eva Goppoldova, Pavla Selepova, MA, Social Psychiatry, Prague Psychiatric Center, Prague, Czech Republic

The SOS is a 10-item scale developed to measure the effectiveness of psychiatric treatments, and measures general psychological functioning. Using standard methodology, we translated the scale into Czech and examined the psychometric properties of the Czech version. 207 in-patients admitted to Prague Psychiatric Center from March to October 2003 were included in the study. All patients completed the SOS at admission and discharge. Clinical data such as disease severity were also collected to assess construct validity. The SOS scale was also administered to 170 persons from the general population. The obtained Cronbach's alpha was .92 and the corrected item-to scale total correlations ranged from .56 to .82. A factor analysis reproduced the structure of the English version with one factor, which accounted for 57 % of the variance. The scale discriminated well between patients and control group with patients scoring significantly lower on all SOS items. The patient sample's admission (mean = 28, SD = 14) and discharge scores (mean = 38; SD=11) were significantly different, indicating that the scale is sensitive to treatment changes. The correlation with disease severity was low at admission (Person correlation = -.126; $p = .073$) and stronger at discharge (Person=-.343; $p = .000$) which might show the effect of insight, especially in patients with psychosis. This hypothesis was supported by the fact that psychotic patients had significantly higher scores at admission (mean = 35; SD = 15) than other patients (mean = 25; SD = 13) while at discharge the score of psychotic patients and those with other diagnoses did not differentiate significantly. The Czech version of the SOS was reliable and valid as a measure of change in patients with a variety of psychiatric conditions. Its reliability and factor structure were similar to the American language version. Further research is necessary to understand why psychotic patients scored higher on the scale at baseline than patients with other mental disorders. Supported by research project CEZ MZ 00000023752.

Abstract 115

HEALTH RELATED QUALITY OF LIFE OF PATIENTS WITH RHEUMATOID ARTHRITIS AND SYSTEMIC LUPUS ERYTHEMATOSUS IN INDIA - A COMPARISON

Nikhil Gupta, MBBS, Gurdes S. Bedi, MBBS, Psychiatry, Sahil Khanna, MBBS, Hemraj Pal, MD, Psychiatry, R M. Pandey, PhD, Biostatistics, Rohini Handa, MD, Medicine, All India Institute of Medical Sciences, New Delhi, India

Purpose Of Research:The aim of this study was to increase an understanding of Health Related Quality of Life (HRQOL) domains affected in Indian patients with Rheumatoid Arthritis (RA) and Systemic Lupus Erythematosus (SLE).
Subject sample and statement of methods:92 patients with RA and 75 patients with SLE were studied at a Rheumatology Clinic in New Delhi, India. The median age was 42.5 yrs (range 20 to 68 yrs) in patients of RA and 35 yrs (range 16 to 68) in patients of SLE. Quality of life was measured using the WHOQOL-BREF questionnaire in both groups. HRQOL scores were adjusted for age and gender imbalance in all domains. Disease activity was measured using DAS28 scores in RA and MEX-SLEDAI scores in SLE patients.
Summary of results:The mean adjusted HRQOL scores of patients with RA in the Physical, Psychological, Social and Environmental domains of the WHOQOL-BREF were 11.96(SD 2.90), 13.01(SD 2.93), 14.47(SD 3.18) and 13.19(SD 2.61) respectively. Corresponding scores of patients with SLE were 12.99(SD2.90), 13.07(SD 2.94), 15.22(SD 3.18) and 14.23(SD 2.61) respectively. This difference in the mean scores was more apparent in patients with greater disease activity. There were significantly lower mean HRQOL scores in RA patients with DAS28 scores greater than 5.1 as compared to SLE patients with MEX-SLEDAI scores greater than 5. The scores of these RA and SLE patients were 9.12 (SD 2.89) and 13.18 (SD 2.95) in the physical domain ($p=0.001$); 14.05 (SD 2.89) and 16.74 (SD 2.95) in the social domain ($p=0.01$); 12.11 (SD 3.15) and 15.40 (SD 3.21) in the environmental domain ($p=0.004$).
Conclusions:Patients with RA have significantly poorer Quality of life scores than those of patients with SLE.

Abstract 116

HEALTH-RELATED QUALITY OF LIFE OF PATIENTS WITH CERVICAL CANCER IN TAIWAN

Hui-Chu Lang, PhD, Maggieleeds Chuang, MA, National Yang-Ming University, Institute of Hospital and Health Care Administration, Taipei, Taiwan

Purpose of Research: Cervical cancer ranks the highest incidence female cancer in Taiwan. Due to the early detection of the mass Pap screening program for all women aged 30 and above since 1995 and advanced medical interventions, the growing survivors live longer than ever before. Combined with the increasing trend of incidence rate and decreasing trend of mortality rate of cervical cancer it indicates that the health-related quality of life of those patients should be an important issue of concern. However, there are still few studies devoted on cervical cancer patients' health-related quality of life. The purpose of the study was to measure the quality of life of cervical cancer patients. **Subject Sample and Statement of Methods:** The study was a cross-sectional study. We used three well-established scales EQ-5D, EORTC QLQ C-30, SF-8 to measure health-related quality of life of cervical cancer patients. The study patients were referred by their doctors before they participated our study at three medical centers. The survey was conducted by face-to-face interview from April 2003 to Oct. 2003. Totally 281 patients were recruited. **Summary of Results:** The mean age of participants was 57.6 years old. Most are married with a lower level of education and income. Regarding EQ-5D, 87.5% of patients reported no problem with mobility; 98.7%, 91.5%, 67.3% and 69.0% of patients reported no problems with self-care, usual activities, pain and discomfort, and anxiety respectively. Most item scores of QLQ-C30 were higher than the reference values, which were provided by the EORTC Group. The physical component score of SF-8 was 46.9 and mental component score was 65.9. The mental component score and bodily pain score were higher than US general population. **Conclusion:** Our results showed that cervical cancer patients in Taiwan enjoy a high quality adjusted life years.

Abstract 117

VALIDATION AND REDUCTION OF FACT/GOG-NTX SUBSCALE FOR PLATINUM-TAXOL INDUCED NEUROLOGICAL SYMPTOMS

Helen Q. Huang, MS, Mark F. Brady, PhD, Statistical and Data Center, Gynecological Oncology Group, Buffalo, NY, David Cella, PhD, Institute for Health Services Research & Policy Stud, Northwestern University Feinberg School of Medicine, Evanston, IL 60201, IL

Purposes: FACT/GOG-NTX subscale was developed to assess patient-reported neurologic symptoms by chemotherapy. This study assesses whether it is reasonable to reduce the subscale while retaining or even improving its psychometric performances. **Methods:** The FACT/GOG-NTX subscale includes 11 items assessing sensory (4 items), motor (3 items), hearing (2 items) symptoms, and possible functional impact (2 items). It was administered to 263 patients with advanced endometrial cancer prior to each of 7 courses of chemotherapy in a Gynecological Oncology Group (GOG) clinical trial. The primary objective of this trial was to assess the impact of adding paclitaxel (TAP, n=134) to standard doxorubicin/cisplatin (AP, n=129) chemotherapy on survival. **Results:** The patient-reported sensory symptom scores (sum of 4 item scores) increased significantly over the treatment duration ($p < 0.001$) in TAP compared to AP. The difference was 0.167 at baseline and was 0.59, 1.60, 2.06, 2.49, 3.38 and 3.18 prior to each of six subsequent assessments, which represented approximately 80% treatment differences in the original 11-items NTX scores. Hearing scores and motor scores were similar in both arms and no obvious incremental trends were observed. The reported functional scores (including trouble buttoning and trouble feeling the shape of small objects) were statistically higher in TAP. However, the maximum treatment differences were 0.42 ($p < 0.01$, at the 4th assessment), which accounted for only 14% of original score difference. Reliability of the subscale was evaluated for patients in TAP arm. At the 4th and 7th assessments, the Cronbach's α s of 11-item subscale were 0.817 and 0.838 and increased to 0.852 and 0.889 with only 4 sensory items. **Conclusion:** 4 sensory symptoms in the FACT/GOG-NTX subscale can be used to assess cisplatin-taxol induced neurologic symptoms in clinical oncology without compromising the psychometric properties. Two functional items may be added if higher doses or prolong treatment is planned.

Abstract 118

CORRELATIONS BETWEEN WHO-DAS II AND QLS IN PATIENTS WITH SCHIZOPHRENIA

Alexandra Christodouloupoulou, MSc, Health Economics Unit, AstraZeneca SA, Athens, Greece, John Yfantopoulos, PhD, Political Science and Public Administration, University of Athens, Athens, Greece, Antonis Karokis, MSc, Health Economics Unit, AstraZeneca SA, Athens, Greece, Vassilia Papagiannopoulou, BA, Political Science and Public Administration, University of Athens, Athens, Greece, Venetsanos Mavreas, MD, PhD, Psychiatry, University of Ioannina, Ioannina, Greece

A shift of interest from symptoms control to the improvement of the quality of life, functioning and disability of patients with schizophrenia is recently observed. This study aims to evaluate the relationship between quality of life and disability, correlating the World Health Organization Disability Assessment Schedule (WHO-DAS II) with a disease-specific questionnaire, the Quality of Life Scale (QLS), in patients with schizophrenia. In an open, 52-week duration study in 8 Greek psychiatric hospitals, 170 patients diagnosed with schizophrenia (DSM-IV criteria), hospitalized or outpatients, newly diagnosed or in acute exacerbation were assessed at baseline and 3, 6, 12 months. Correlations analysis was used to identify the relationship between WHO-DAS II and QLS sub domains scores. Baseline and 6-month follow-up data are presented in this analysis. The statistically significant correlation between WHO-DAS II Relations with others with QLS Interpersonal relations and Intrapsychic foundations domains at baseline became stronger at 6-months ($r=-0.309, -0.269$ vs. $-0.479, -0.384, p<0.01$). WHO-DAS II Social participation was significantly correlated with QLS Interpersonal relations and Intrapsychic foundations domains only after 6-months follow-up ($r=-0.055, -0.084$ vs. $-0.322, -0.244, p<0.01$). The baseline significant correlation WHO-DAS II Activities of daily life and QLS Common objects and activities was expectedly not significant at 6 months ($r=-0.496$ ($p<0.01$) vs. -0.120). Quality of life in schizophrenia is correlated with patients' functioning and disability in some but not all domains. This is due to the fact that disability and quality of life represent overlapping but not identical concepts. Both types of instruments, together with psychopathology measures, should be used for a comprehensive assessment of the health state of patients with schizophrenia.

Abstract 119

CORRELATIONS BETWEEN WHO-DAS II AND EQ-5D AND SF-36 IN PATIENTS WITH SCHIZOPHRENIA

Alexandra Christodouloupoulou, MSc, Health Economics Unit, AstraZeneca SA, Athens, Athens, Greece, John Yfantopoulos, PhD, Political Science and Public Administration, University of Athens, Athens, Greece, Antonis Karokis, MSc, Health Economics Unit, AstraZeneca SA, Athens, Greece, Vassilia Papagiannopoulou, BA, Political Science and Public Administration, University of Athens, Athens, Greece, Venetsanos Mavreas, MD, PhD, Psychiatry, University of Ioannina, Ioannina, Greece

Quality of life measurement in schizophrenia includes subjective and objective assessment of the disease and its treatment impact on patients' well-being, disability and functioning. The study aims to examine correlations between general health profiles EuroQol-5D (EQ-5D) and Short Form-36 (SF-36) and the World Health Organization Disability Assessment Schedule (WHO-DAS II) in patients with schizophrenia. In an open, 52-week duration study conducted in 8 Greek psychiatric hospitals, 170 patients diagnosed with schizophrenia (DSM-IV criteria), hospitalized or outpatients, newly diagnosed or in acute exacerbation were assessed at baseline and 3, 6, 12 months. Outcome measures included generic and specific QoL indices and psychopathology estimates. EQ-5D and SF-36 were self-assessed by patients. WHO-DAS II was assessed by the investigator. Correlations between the generic QoL indices at baseline and 6-month reported data are presented. Baseline and 6-month significant correlations are observed between: SF-36 Social functioning-WHO-DAS II Relations with others, and Social participation; SF-36 Physical functioning-WHO-DAS II Mobility in space; SF-36 Role physical-WHO-DAS II Activities of daily life; EQ-5D Mobility, and Self-care-WHO-DAS II Mobility in space, Understanding & communication and Activities of daily life. Baseline but not 6-month significant correlations exist between: SF-36 Role physical-WHO-DAS II Mobility in space; SF-36 Physical functioning-WHO-DAS II Activities of daily life; EQ-5D Usual activities and Pain/Discomfort-WHO-DAS II Activities of daily life. The study reveals a statistically significant correlation between the general health profiles and the disability index. Both types of outcome indicators may be used to provide for a more comprehensive evaluation of schizophrenia patients' quality life as they combine objective and subjective assessments.

Abstract 120

VALIDITY OF THE PATIENT PERCEIVED QUALITY OF TREATMENT (PPQT) TO ASSESS THE SATISFACTION, PREFERENCE AND WILLINGNESS OF OAB PATIENTS TO REUSE TREATMENT

J.R. Landis, PhD, Biostatistics and Epidemiology, University Pennsylvania School of Medicine, Philadelphia, PA, K Kahler, PhD, Health Economics & Outcomes Research, Novartis Pharma, East Hanover, NJ

This study assessed convergent and discriminant validity of the 3-question PPQT in a multicenter, randomized, double-blind, placebo-controlled, parallel-group study of darifenacin in patients with overactive bladder (OAB). The PPQT was administered at study end as 3 complementary patient-reported outcome (PRO) questions scored on a 5-point Likert scale, assessing patients' satisfaction with treatment (PS), global preference for treatment (PP) and willingness to reuse treatment (PWTRT). The discriminant validity of the questions was assessed by comparisons with the study's main efficacy endpoints (incontinence episodes, micturitions and episodes of urgency). Convergent validity (strength of association between 2 methods) was evaluated with the Spearman's rank correlation for the PS, PP, and PWTRT with the change between randomization and final visits in the mean values of all domains of the King's Health Questionnaire (KHQ). After washout and 2-week placebo run-in, OAB patients (21-88y) were randomized to darifenacin controlled-release tablets (n=330) or matching placebo (n=109) for 12 weeks. The degree of association between the 3 PPQT questions was confirmed by the observed correlation coefficients between patient responses to each item of the PPQT (0.74 for PP and PWTRT, 0.68 for PS and PWTRT, and 0.69 for PS and PP). For discriminant validity, a strong relationship was observed between the main efficacy endpoints and PP, PS, and PWTRT; PP, PS and PWTRT increased with increases in treatment effects on clinical outcomes. Spearman's rank correlations with the PP, PS, and PWTRT were generally 0.3 or 0.4 with the KHQ dimensions, indicating neither lack of association nor redundancy. This psychometric validation of the PPQT shows that the PP, PS and PWTRT are sufficiently related, yet each measure a different concept and have a high degree of validity, confirming their relevance for clinical use. Supported by Novartis Pharma, Switzerland

Abstract 121

VISUAL DISABILITIES AND QUALITY OF LIFE IN GLAUCOMA

L. Jay Katz, MD, Glaucoma Service, Wills Eye Hospital, Philadelphia, PA, John Walt, MBA, Global Outcomes Research, Allergan, Inc., Irvine, CA

The purpose of this presentation is to evaluate the clinical evidence that quality of vision and quality of life (QOL) are impaired in glaucoma patients long before they progress to blindness. A MEDLINE search identified studies of the quality of vision and QOL in glaucoma. Results of the studies were reviewed and compared. Studies using vision-targeted QOL surveys have shown that QOL is impaired in glaucoma. Patients with mild field loss have diminished QOL compared with control subjects. QOL is further decreased as visual field loss becomes more severe. Patient performance on tasks involving light and dark adaptation, glare, and peripheral vision can be impaired in early to moderate glaucoma. Everyday tasks such as walking without bumping into or tripping over objects, reading a newspaper, and crossing a road become increasingly difficult as glaucoma progresses. Glaucoma has been identified as a risk factor for car crashes, and patients are often particularly concerned about their difficulties with driving. Many patients express fear of becoming blind and dependent on others. A utility analysis reported that glaucoma patients on average would give up 6.1% of their remaining lifespan to have perfect vision. In summary, even mild visual field loss is associated with visual disabilities that affect the daily lives of glaucoma patients. Patient QOL is decreased further as the disease progresses. The most important goal of treatment is to preserve patient QOL. Early, aggressive treatment is desirable to prevent visual disabilities and preserve QOL in glaucoma.

Abstract 122

QUALITY OF LIFE (QOL) VARIABLES INFLUENCE SURVIVAL IN RADIATION THERAPY ONCOLOGY GROUP (RTOG) HEAD & NECK TRIALS

Benjamin Movsas, MD, Andre Konski, MD, RadOnc, Fox Chase Cancer Center, Phila, PA, Jonathan Harris, MS, Statistics, RTOG, Phila, PA, James Coyne, PhD, Abramson Cancer Ctr, Univ of Pennsylvania, Phila, PA, Clement Gwede, PhD, RadOnc, H.Lee Moffitt Cancer Ctr, Tampa, FL, Adam Garden, MD, RadOnc, M.D. Anderson Cancer Ctr, Houston, TX, Sharon Spencer, MD, RadOnc, U of Alabama Med Ctr, Birmingham, AL, Christopher MD, MD, RadOnc, Radiation Oncology Ctr, Sacramento, CA, Deborah Watkins-Bruner, PhD, Population Science, Fox Chase Cancer Center, Phila, PA

The purpose of this study was to evaluate the influence of QOL (via FACT-G, version 2.0) in addition to disease & socio-demographic factors on the survival of patients (pts) treated on RTOG head & neck cancer clinical trials. 1510 pts treated on RTOG 9003 (locally advanced disease) and RTOG 9111 (laryngeal preservation) were potentially analyzable; baseline FACT-G data were available for 1093. Patients with and without FACT-G had different survival rates for RTOG 9111 (HR=1.423; p=0.025), but not for RTOG 9003 (HR=0.972; p=0.73). The mean FACT-G and H&N subscale scores were 86.4 and 21.2 with a standard error of measurement (SEM) of 5.3 and 4.3, respectively. A Cox model stratified by protocol assessed which QOL measures (baseline FACT-G & its subscales) had prognostic impact after accounting for tumor and socio-demographic variables, including marital status, # in household, education level, household income, & cigarette use. The QOL measures were considered as continuous variables with higher scores associated with better QOL. On univariate analysis, FACT-G & all the subscales (except for doctor score and emotional well-being) showed significantly better survival for higher values. On multivariate analysis, T- & N-Stage & KPS were significant. Socio-demographic variables found to have independent prognostic value were smoking status (HR=1.512; p<0.0001), age (HR=1.020; p<0.0001), race (HR=1.288; p=0.0061), & marital status (HR=1.277; p=0.0032). No single QOL variable was found to be prognostically significant when added to the fitted model. However, when two subscales were simultaneously added to the model, functional (HR=0.980; p=0.0056) and emotional (HR=1.029; p=0.0191) well-being were found to be significant. Unlike the functional well-being scale, lower emotional well-being scores were unexpectedly associated with better survival. In conclusion, some QOL & socio-demographic variables do have independent prognostic value for survival, which may identify patients requiring additional support & early interventions to improve outcome.

Abstract 123

VALIDATION OF THE EUROQOL THERMOMETER IN PATIENTS WITH CHRONIC OBSTRUCTIVE PULMONARY DISEASE

Josephine Norquist, MS, Carla Demuro-Mercon, MA, Linda Nelsen, MPH, Jennifer Turpin, Nancy Santanello, Epidemiology, Merck & Co., Blue Bell, PA

Objective: The EuroQoL Visual Analog Scale (EQ VAS) is a generic health-related quality of life measure. Limited data are available on its measurement properties among patients with Chronic Obstructive Pulmonary Disease (COPD). **Methods:** Patients with COPD were enrolled in a 4-week, multicenter, observational study and completed the EQ VAS, SF-36, Chronic Respiratory Questionnaire (CRQ), St. George's Respiratory Questionnaire (SGRQ), Baseline Dyspnea Index (BDI), Transition Dyspnea Index (TDI) and a COPD Symptoms Diary. Spirometry was performed at baseline, week 2 and week 4. Patient were categorized into three groups according to physicians' assessment of baseline COPD status: stable (no treatment change), unstable (requiring additional medications for symptoms) and exacerbated (need for antibiotics or oral corticosteroids). At week 4, patients completed a global evaluation of change in COPD status and were grouped as worse, unchanged, or better. **Results:** A total of 109 patients were enrolled in the study (stable n = 59, unstable n = 21, exacerbated n = 29). Baseline EQ VAS scores were significantly different between stable, unstable and exacerbated groups (p < 0.05). As hypothesized, EQ VAS scores at baseline were moderately to strongly correlated with BDI, CRQ and SF-36 dimension scores (range 0.33 to 0.82, p < 0.001) and with COPD Symptoms Diary, overall physicians' rating of symptoms, SGRQ dimensions and SOBQ score (range -0.38 to -0.69, p < 0.001). Moderate correlations were obtained with baseline b-agonist use (r = -0.27, p<0.05); weak correlations were obtained with variables assessing lung function. With repeated measurement, the EQ VAS was reliable in stable patients (intraclass correlation coefficient = 0.78). At week 4 patients who rated themselves as being worse had an EQ VAS mean change score of -2.85 (SD 15.47) while patients rating themselves as unchanged or better had mean change scores of 3.25 (SD 15.22) and 7.14 (SD 18.05), respectively. **Conclusions:** The results from this study suggest that the EQ VAS is valid and responsive for use in clinical research in patients with COPD.

Abstract 124

IS THE GASTROINTESTINAL SYMPTOM RATING SCALE (GSRS) A USEFUL TOOL FOR MEASURING THE GI EFFECTS OF NSAID TREATMENT?

Nancy Kline Leidy, PhD, Health Economics & Outcomes Research, MEDTAP International, Inc., Bethesda, MD, Liberty Fajutrao, Jonas Carlsson, PhD, Health Economics & Outcomes Research, AstraZeneca R&D, Södertälje, Sweden, Ingela Wiklund, PhD, Outcomes Research, AstraZeneca R&D, Molndal, Sweden

It is well-known that non-steroidal anti-inflammatory drugs (NSAIDs) relieve the pain and stiffness of arthritis. Unfortunately, this relief is often accompanied by distressing gastrointestinal (GI) side effects. The purpose of this study was to evaluate the suitability of the Gastrointestinal Symptom Rating Scale (GSRS), a validated measure of symptom distress in the GI population, for assessing GI distress and outcomes of treatment in patients treated with NSAIDs. Secondary analyses were performed on baseline data from three randomized controlled trials (N=1,601). End-of-study values were used to evaluate the measure's responsiveness to change. The factor structure of the GSRS was consistent with that found in the GI population. The instrument also exhibited evidence of reliability, with internal consistency levels averaging 0.76 across the 5 subscales comprising the measure and ranging from 0.59 (Abdominal Pain) to 0.84 (Diarrhea). Subscales correlated moderately with investigator rating of similar symptoms and the Quality of Life in Reflux and Dyspepsia (QOLRAD) scale, a condition-specific measure of the impact of GI symptoms on health-related quality of life (HRQL). Correlations with a generic measure of HRQL, the SF-36, were weak to moderate. Effect size (ES) estimates for patients reporting improvements in their GI symptoms during the course of the trials ranged from 0.28 (Diarrhea) to 1.13 and 1.20 for the Reflux and Abdominal Pain scales, respectively. Together, these results suggest that the GSRS is a suitable instrument for assessing symptom distress and evaluating GI symptom outcomes in patients on NSAIDs therapy.

Abstract 125

THE USE OF COMPLEMENTARY & ALTERNATIVE MEDICINE TO IMPROVE QUALITY OF LIFE IN THE SYMPTOM-MANAGEMENT TRIALS OF THE NCI- SUPPORTED COMMUNITY CLINICAL ONCOLOGY PROGRAM

David R. Buchanan, DrPH, Ann O'Mara, PhD, Joseph Kelaghan, MD, Lori Minasian, MD, Community Oncology and Prevention Trials Research Gr, National Cancer Institute, Bethesda, MD

Approximately 40% of the US adult population is estimated to use complementary & alternative medicine (CAM), with higher prevalence among white and well-educated patients. To provide sound scientific information on the safety and efficacy of their use, the US Congress established the National Center for Complementary & Alternative Medicine in 1998. The National Cancer Institute has provided support for 20 Phase III RCTs of CAM for symptom management and QOL improvement conducted by the Community Clinical Oncology Program. Intervention modalities include botanical agents (St. John's Wort, ginger, ginkgo biloba, soy, etc.), vitamins, acupuncture, and guided meditation for stress management. The inclusion of general QOL measures is more common in CAM studies than in conventional pharmacological symptom management trials. This study presents a summary of the research design issues that are unique to the use of CAM in QOL/cancer symptom-management trials. The data are based on an extraction of the scientific review committees' comments on these trials from the NCI archives. External research design issues unique to the use of CAM include: sample contamination (drop-ins, crossovers); the form of agent (pill, powder, extract, etc.); type of agent (e.g., different species of ginger); quality control over dose (amount of active agent in commercially available products); product adulteration; potential interactions with other drugs; methods of procurement; perceptions of side-effects in different populations (e.g., marinol); and determinations of appropriate dose levels (bounds of clinically sufficient as well as clinically excessive). Internal research design issues include accrual rates, the identification of appropriate control groups and the use of QOL measures unique to CAM studies (e.g., spiritual integration).

Abstract 126

PSYCHOMETRIC CHARACTERISTICS OF THE FUNCTIONAL ASSESSMENT OF CANCER THERAPY-GENERAL VERSION 4 (FACT-G) IN PATIENTS WITH METASTATIC CANCER (MC) OR HEMATOLOGIC MALIGNANCIES (HM)

Susan F. Marden, PhD, Nursing and Patient Care Services, Sharon A. Mavroukakis, MS, National Cancer Institute, Margaret Bevans, MSN, Nursing and Patient Care Services, National Institutes of Health, Bethesda, MD, Karen Soeken, PhD, School of Nursing, University of Maryland, Baltimore, MD, Donald E. White, MS, Tatiana Beresneva, MD, National Cancer Institute, Paula Muehlbauer, MSN, Nursing and Patient Care Services, National Institutes of Health, Bethesda, MD, Nancy Kline Leidy, PhD, Center for Health Outcomes Research, MEDTAP International, Bethesda, MD, U.S.A

The FACT-G is a cancer-specific health-related quality of life (HRQL) instrument widely used across cancer populations. We examined the psychometric characteristics of the FACT-G in adult patients with MC (n =212) or HM (n=73) with good performance status (PS). Patients (N=285) enrolled in prospective, phase II clinical trials and associated HRQL studies completed a self-administered survey prior to investigational therapy. HRQL measures included the 1) FACT-G and 2) Short Form-36 Survey (n=192; SF-36). The Symptom Distress Scale (n=166; SDS), a measure of the degree of cancer symptom discomfort was also completed, with higher scores indicating a greater level of distress. Descriptive statistics summarized FACT-G subscale (physical (P), social (S), emotional (E), functional (F)) and total scores. Internal consistency reliability was estimated using coefficient alpha (Cronbach formula). Construct validity was assessed by (1) examining relationships between FACT-G scores and SF-36 and SDS scores using Pearson correlations and (2) determining whether FACT-G scores differed between MC and HM patients using multivariate analyses of variance. Most patients were male (60%), Caucasian (81%) or Hispanic (10%), on average 47 +/- 12 years of age with ECOG status 0 (86%). MC patients had metastases to the peritoneum or liver (56%) or stage IV melanoma (44%). HM patients had acute (18%) or chronic (40%) leukemia, lymphoma/multiple myeloma (30%) or myelodysplastic syndrome (12%). FACT-G mean scores ranged from 59-75 (Table). Coefficient alpha estimates were >/.70 for FACT-G scores except S scores in HM group (Table). FACT-G scores were positively correlated with SF-36 scores (r = .17 to .69; p<.05). Negative correlations were found between FACT-G and SDS scores (r = -.30 to -.78; p<.01). HRQL of groups differed (Pillai's Trace =.071, F(5,279)= 4.2, p=.001) with HM group scoring higher on E (p=.006). Results suggest FACT-G is reliable and valid for use in cancer patients with MC or HM with good PS prior to investigational therapy.

Scal	Total Sample (N=285)	Metastatic (MC) (n=212)	Hematologic (HM) (n=73)
FACT-G Physical: Mean (SD)	74(18)	75(19)	73(15)
Coefficient alpha^	.88	.89	.81
FACT-G Social: Mean (SD)	71(17)	73(17)	67(16)
Coefficient alpha^	.76	.79	.64
FACT-G Emotional: Mean (SD)	63(15)	62(14)	67(17)
Coefficient alpha^	.78	.80	.70
FACT-G Functional Mean (SD)	63(17)	64(17)	62(17)
Coefficient alpha^	.88	.88	.88
FACT-G Total : Mean (SD)	60(8)	60(8)	59(8)
Coefficient alpha^	.91	.92	.88
Scored 0-100			
Higher scores=better HRQL			
^Cronbach formula			

Abstract 127

IS IT POSSIBLE TO PREDICT PATIENTS' QOL AFTER OPEN HEART SURGERY?

Zoran R. Potic, BSE, Djordje Jakovljevic, Prof, Marijela M. Potic, BSE, Ninoslav D. Radovanovic, Prof, Clinic of Cardiovascular Surgery, Institute for Cardiovascular Diseases, Sremska Kamenica, Serbia & Montenegro

Using QOLi-NS index, an integral numerical value for measuring disease specific health related Quality of life in cardiovascular patients undergoing open heart surgery, we tried to develop a simple (additive) model for predicting improvement of QOL after surgery. Target variable was improvement of QOL six months after open heart surgery calculated as difference between QOLi-NS scores before and after surgery. For factors influencing changes in QOLi-NS we chose widely available preoperative clinical data. We collected relevant data for 294 consecutive patients (both risk factors and QOLi-NS scores before and six months after operation). After applying multivariate stepwise linear regression using QOLi-NS improvement as dependent variable and set of preoperative risk factors as independent ones, a simple linear (additive) model with 18 risk factors was created. Correlation between predicted QOL improvement and observed was very strong ($r=0.78$ $p<0.001$). Patients from training set were stratified into four groups according to the predicted QOLi-NS improvement: The developed method was applied to another group of 465 patients and also has shown strong capabilities of prediction. Obtained method may be used for prediction of patient's outcome after open heart surgery in order to help patients to weight risks and benefits of an operation and as quality monitoring tool for healthcare professionals.

Abstract 128

QUALITY OF LIFE AND LONG TERM SURVIVAL AFTER OPEN HEART SURGERY

Zoran R. Potic, BSE, Djordje Jakovljevic, Prof, Marijela M. Potic, BSE, Bogoljub Mihajlovic, Prof, Ninoslav D. Radovanovic, Prof, Cardiovascular Surgery, Institute for Cardiovascular Diseases, Sremska Kamenica, Serbia & Montenegro

The aim of this study was to estimate long term benefit of open heart surgery in seven year period, using quality of life and cumulative survival as outcomes. For assessment of quality of life, we used diseases and treatment specific health related quality of life questionnaire and derived integral numeric QOL index QOLi-NS (0-100). Zero value represents complete inability for any functioning and value 100 means perfect health. The cumulative survival rate was calculated using actuarial method. The prospective study involved 787 consecutive patients. The survival adjusted QOL (QOLadj) was calculated according the following formulae $QOLadj(t) = QOLi-NS(t) * Survival(t)$. The benefit of operation was calculated as difference between postoperative survival adjusted QOL and preoperative QOL (QOLi-NS= 40.9). Postoperative results are presented as: Average QOLi-NS, QOLadj(t), QOL benefit and survival. Patients' benefit, concerning survival adjusted QOL, is the greatest at one year after the surgery and declines during the follow-up period, but is very clinically and statistically significant. The mean QOL benefit in seven year follow-up period is 29. The survival adjusted QOL could be the most appropriate measure for the evaluation of open heart surgery outcome.

Abstract 129

VALIDATION OF THE OTITIS PARENT QUESTIONNAIRE (OPQ)

Jordana K. Schmier, MA, Health, Exponent, Alexandria, VA, Mary A. Cifaldi, PhD, Center for Pharmaceutical Appraisals & Outcomes Res., Abbott Laboratories, Abbott Park, IL, Michael T. Halpern, MD, PhD, Health, Exponent, Alexandria, VA

Introduction: The study objective was to validate the Otitis Parent Questionnaire (OPQ), a parent-completed instrument assessing satisfaction and related outcomes for acute otitis media (AOM) in pediatric populations, by evaluating its relationship with clinical and other outcomes measures. **Methods:** The OPQ was used in a multi-center randomized clinical trial of two AOM therapies. The OPQ consists of 15 questions (11 items) and was administered at visit 3 (day 12-14). Available clinical measures included physician-rated clinical efficacy (visits 2-4) and symptom assessments (visits 1-3). Analyses included scale development and scoring, internal consistency reliability, and construct validity. Analyses were conducted across treatment assignment. Significance was declared at $p < 0.05$. **Results:** A total of 367 parents completed the OPQ. Four scales plus one independent item were identified: Overall Impression, 4 items on satisfaction, ease of drug use, taste, and likelihood to use again; Productivity, 2 items on work loss and missed day care; Lower GI Side Effects, 2 items on diarrhea and diaper rash; Compliance, 2 items; and a separate item on vomiting. Cronbach's alpha was > 0.60 for all scales except Compliance. Overall Impression was significantly positively related to clinical efficacy at all visits and significantly negatively related to symptoms, including ear pain, mobility, fullness, irritability, and erythema. Lower GI and Productivity were significantly related to erythema and fullness at end of treatment but not to clinical efficacy. There were significant differences in mean Overall Impression scores by clinical outcome (failure/success). No overall score is recommended as inter-scale correlations are not significant; each scale presents a unique construct. **Conclusions:** The OPQ is a valid and useful tool for assessing parent evaluations of treatment for AOM. The various scales measure different outcomes, not all of which are related to clinical efficacy.

Abstract 130

COPD PATIENTS' BURDEN OF EXACERBATIONS

Elisabeth S. Ståhl, PhD, Clinical Science, AstraZeneca R&D, Lund, Sweden, Claes-Göran Löfdahl, MD, PhD, Respiratory Medicine, University Hospital, Lund, Sweden

Aim: To understand the impact and burden of COPD exacerbations and how patients are affected. **Methods:** 26 patients (aged >40 years, 77% male) with COPD (GOLD stages III-IV, $n=12$; stages I-II, $n=14$) from France, Germany and the UK were asked about periods of worsening symptoms that led to an exacerbation requiring medical intervention: Q1) How often do you experience worsening symptoms (more difficult to breathe, more coughing or greater chest tightness, compared with one of your 'good days') and need to take oral steroids/antibiotics and/or have to go to hospital? Q2) How burdensome/troublesome/hard/difficult are periods of increased symptoms? Q3) In what way are they burdensome/troublesome/hard/difficult? Answers were recorded. **Results:** A1) Patient responses for exacerbation frequencies were grouped as " >1 /month" ($n=9$), "1-2/month" ($n=9$), and "from 1-3/year to almost never" ($n=8$). Almost all patients (24/26) with COPD had periods of worsening symptoms leading to exacerbations 1-3 times/year. A2) 9 patients were depressed and breathless at rest, and said these periods were "absolutely awful"; 12 patients were tired and breathless with discomfort; 5 patients had unpleasant breathing problems. A3) 15 patients "wanted to give up", felt panic, and had no social life; 9 patients found walking difficult, and increased their reliever medication use; 2 patients "tried to keep on the move". **Conclusion:** Most patients with COPD experience several exacerbations each year, which restrict daily activities and have a negative effect on their wellbeing.

Abstract 131

THE EFFECTS OF SAMPLE BIAS ON THE FACTOR STRUCTURES OF QUALITY OF LIFE INSTRUMENT

Richard D. Lennox, PhD, The Health Statistics Lab, Psychometric Technologies, Hillsborough, NC, Mark Atkinson, PhD, Worldwide Outcomes Research, Pfizer, Inc., San Diego, CA

The biased selection of patients with moderate-to-high level of disease severity for most clinical trials has long been a source of concern to PRO instrument developers, who rely on more general samples to validate new instrumentation. Despite this concern, few studies have attempted to systematically evaluate consider the issue. This study attempts to provide some insight into the extent that selection bias in the study of endometriosis results in discernable differences in latent PRO constructs between high and low disease severity on some common quality of life and clinical measures. A detailed analysis was used to examine whether illness severity resulted in: 1) a different pattern of non-trivial factors, 2) significantly different factor loadings, 3) significantly difference inter-factor correlations, or 4) a different patterns of salient item loadings. Using data from a clinical trail for treatment of endometriosis, the study compared factor structure from subsamples of high and low disease severity using the SF36, the Endometriosis Scale and Kupperman Scale. Exploratory factor analysis was used to evaluate the factor structure of the total sample vs. severity subgroups. A multi-sample confirmatory factor analysis of the three measures across the high and low severity dichotomy revealed the two groups to be invariant across the number of non-trivial factors. Comparison of the nested X², goodness-of-fit criteria and the Lagrange multipliers for the constrained models found no significant differences across the groups in terms of standardized loadings and intercorrelations among the factors. Comparison of standardized loading on the unconstrained models found that all three instruments yielded identical patterns of loading above .40 across the high and low disease severity groups. The results suggest that the concerns regarding the generalizability of QOL instruments, developed using convenience sampling, for application in clinical trials may be unfounded.

Abstract 132

DEVELOPMENT AND PRELIMINARY PSYCHOMETRIC EVALUATION OF THE MOTIVATION AND ENERGY INVENTORY - SHORT FORM (MEI-SF)

Sheri E. Fehnel, PhD, Patient-Reported Outcomes, RTI Health Solutions, Research Triangle Park, NC, Heather M. Edin, MSPH, Global Health Outcomes North America, GlaxoSmithKline, Research Triangle Park, NC, Lori D. McLeod, PhD, Patient-Reported Outcomes, RTI Health Solutions, Research Triangle Park, NC, Susan L. Hogue, PharmD, Global Health Outcomes North America, GlaxoSmithKline, Research Triangle Park, NC

OBJECTIVE: The Motivation and Energy Inventory (MEI), a 27-item questionnaire with a 4-week reference period, was developed to facilitate the identification of treatments effective in restoring the vitality of depressed patients. The current project was undertaken to create an acute, short form of the MEI, capable of both reducing patient burden and maximizing responsiveness to treatment effects. **METHODS:** Data from two clinical trials and a separate psychometric evaluation of the MEI were utilized. To inform item reduction, three steps were taken. First, a series of pairwise t-tests was conducted to compare changes on each MEI item for responders (subjects whose HAM-D scores decreased by 50% or more) and non-responders, with the goal of identifying items not optimally sensitive to differences between these groups. Second, item-level changes were correlated with changes in HAM-D scores for both groups, as well as the entire sample, to identify items only weakly related to changes in symptom severity. Third, inter-item correlations were examined to locate potentially redundant item pairs. The psychometric properties of the final item set (as administered within the context of the MEI) were then investigated using standard techniques. **RESULTS:** Responsiveness analyses resulted in the removal of four items addressing behaviors (rather than feelings), such as participating in social activities and exercise. The examination of highly correlated item pairs resulted in the removal of five additional items. Using data from only the remaining 18 items, internal consistency, test-retest reliability, validity and responsiveness estimates obtained were comparable to those of the full instrument. **CONCLUSIONS:** While the psychometric properties of the MEI-SF as a stand-alone instrument, utilizing a one-week reference period, have yet to be established, the methods used to choose its items and the results of preliminary analyses suggest that this instrument may be of even greater utility than the original MEI, particularly for trials of short duration.

Abstract 133

QUALITY OF LIFE AMONG LITHUANIAN UNIVERSITY STUDENTS AND TENDENCIES TO SUICIDE Danute Ducinskiene, MD, Social Medicine, Kaunas University of Medicine, Kaunas, Lithuania

The aim of the study was to evaluate quality of life (QoL) among university students using the WHOQoL-BREF instrument and to explore tendencies of suicides among students in relation with their quality of life. 919 third year students from three universities (Kaunas University of Medicine, Kaunas University of Technology and Vilnius University) were involved in the study. The average overall QoL score was 13.71 (from 20 possible). A strong correlation between the physical and the environmental domains ($r = 0.52$), social relations and the psychological domain ($r = 0.5$), as well as between health conditions perceived by students and their mean QoL scores was observed. Differences by gender in the physical domain were also found to be significantly different: female scores were lower than male scores ($p < 0.05$). More than a half of students (50.6%) had a tendency of suicide. The mean of quality of life scores of these students were lower in all domains. The chance of higher social and physical quality of life who never thought about suicide is considerably higher 2.1 and 1.4 times accordingly.

Abstract 134

QUALITY OF LIFE AND HEALTH STATUS OF SEAFARERS AND THEIR FAMILIES IN INDIA

Arti L. Barnes, MB, B.S, Public Health, Blanche L. Barnes., Psychology, Dr. Leo Barnes Foundation, Mumbai, India, Aravind Sugumar, MD, Clinical Research, Zakim Center for Integrative Therapies, Dana Farber Cancer Institute, Boston, MA

Quality of life assessment scale was specifically adapted to seafaring population of India. The final analysis was done on incidentally selected 283 fully completed data forms of seafarers and their families. The seafaring profession is constantly subjected stress due to unique nature of their jobs. A pool of items was taken up to formulate a rating scale. It was subjected to necessary statistical rigor to establish its validity (consensus of experts) and the reliability. The finalized version consisted of 40 items rated on a five point rating scale, evaluating QOL from low to high (higher the score better the QOL). The reliability for internal consistency by Cronbach's Alpha was 0.85. The factor analysis has explained 53% of variance, with nine factors, which reflect different domains of QOL, namely, emotional, QOL-general, Life satisfaction, Physical health, Self esteem, interpersonal interaction, social Milieu, personal needs and others. The predictive ability was tested on Logistic Regression analysis, which showed 83% accurate classification between high and low QOL score. Even, Discriminant analysis correctly classified 71% of the population. The analysis of results has delineated some significant variables that predict the chances of having either high or low QOL. The highly significant but with low QOL were respiratory, orthopedic, cardiac and organ related diseases. However, the diabetes group manifested significantly high QOL. The Life concerns significantly affecting QOL were Loss of work and Physical health status. The demographic variables that negatively impacted on QOL was for retirees and widows, whereas, higher QOL was observed with males, employed, better educated, unmarried, more living space, and number of gadgets available in the house than their cohort groups. The region wise analysis pointed out that seafaring population of Kerala had the highest QOL, followed by Gujarat, Tamil Nadu and then Maharashtra state.

Abstract 135

QUANTIFYING QUALITY: CAPTURING THE WHOLE OF LIFE

Paul O. Bramston, PhD, Grace M. Pretty, PhD, Psychology, University of Southern Queensland, Toowoomba, Queensland, Australia

Subjective quality of life is often quantified by researchers and practitioners as a single figure. Quality of health related life, for example, is often reduced to a total quality of life index that purports to represent an individual's sense of well-being or life satisfaction. While the convenience of a single index is undeniable, the complexity of a multifaceted psychological construct such as quality of life makes it particularly difficult for a single figure to validly represent whole of life. The search for a functional list of domains that comprise life quality has been long and a group of researchers in Australia (Cummins & colleagues) has used meta-analytic techniques to support seven key areas of life as tapping its most important elements. Data will be presented that show that the following

domains provide a reasonably comprehensive coverage of whole of life: standard of living, health, achievements in life, personal relationships, safety, community connections and emotional well-being. These domains have been incorporated into a self-report quality of life measure, ComQol that produces scores for each domain. Data from large population studies will be presented as normative data for each domain and in support of a theoretical mean or gold standard of general quality of life for these domains. Finally, the results of a study using subjective quality of life ratings of 64 volunteer Landcare workers are used to highlight how reductionist a single quality of life index can be, the benefits to researchers and practitioners of the seven domain ratings for different aspects of life and the benefits of using the theoretical gold standard to judge life satisfaction. Voluntary environmental stewardship networks exist across the USA and many other countries in the world and in Australia there is a National Landcare Program that is popularly seen as a successful attempt to address rural environmental protection. We examined how such selfless contributions of time and labour on behalf of the environment are related to life domains.

Abstract 136

QUALITY OF LIFE ASSESSMENT IN AN OUTCOMES TRIAL: USING THE INTERNET TO IMPROVE SELF-MANAGEMENT BY PATIENTS WITH CHRONIC ILLNESS

Deborah M. Miller, PhD, Neurology, Cleveland Clinic Foundation, Cleveland, OH

Purpose: The Mellen Center for Multiple Sclerosis received NIH funding to conduct a randomized trial using the Internet to improve self-management of MS patients. We describe how QOL assessment is key to this study. The study is to develop and test a telehealth system, Mellen Center Care On Line (MCCO) that includes: 1) asynchronous electronic messaging between subjects and MS specialists, 2) subject self-assessment of MS status, monitor change and decide if intervention is indicated, 3) a method for subjects to prepare for upcoming appointments. Study subjects will have all three components; comparison subjects will access only the messaging. The study is based on two conceptual models, Wagner's Model for Effective Chronic Illness Care and the Institute of Medicine's (IOM) guide to assessing telecommunications in health care. Three QOL measures operationalize aspects these models in the trial; The Multiple Sclerosis Quality of Life Inventory (MSQLI), (includes the SF-36 and 9 disease scales), the Sickness Impact Profile (SIP), and the EuroQol. **Methods:** This study offers Wagner's recommended patient self-management support with the MSQLI. With Internet access, subjects will complete the MSQLI every 3 months and, if they wish, more often. The system will show current scale scores and compare to previous scores. Subjects will decide if they find clinically important change and receive support in decision making. Two of the IOM telehealth outcomes evaluation criteria include quality and cost assessment. Our quality assessment will include the SIP and cost effectiveness will be assessed using the EuroQol utility measure. **Summary:** This study includes one QOL measure, the MSQLI, as a process measure and two others, the SIP and the EuroQol as outcome measures. In addition to increasing the profile of QOL studies in NIH funded outcome studies, this investigation will provide a longitudinal study of patient assessed clinically important change and an opportunity to compare the relative performance of psychometric and econometric measures over time.

Abstract 137

USING ITEM RESPONSE THEORY TO DEVELOP AN EFFICIENT AND INFORMATIVE SHORT FORM OF THE CES-D

Timothy E. Stump, MA, Colleen A. McHorney, PhD, Regenstrief Institute, Regenstrief Institute, Indianapolis, IN

The purpose of this poster is to illustrate how item response theory (IRT) can be used to create an efficient short form of the Center for Epidemiological Studies-Depression (CES-D) scale. The CES-D is one of the most widely-used measures of depressive symptoms. Previous CES-D short form (SF) measures have been developed largely using elderly populations and classical test theory. We studied the 20 polytomous CES-D items in the NHANES-I, a national probability sample of 3,059 persons aged 25-74. Using MULTILOG, we obtained item difficulty and discrimination estimates for Samejima's graded response model and generated item characteristic, item information, and test information curves. To create a SF measure based on IRT information, we blinded ourselves to item content. Then, we eliminated items with low discrimination parameters and, thus, low item information. We also eliminated items where there was substantial overlap between option response categories. The resulting CES-D SF had 13 items. To reduce the number of items to half the original scale ($k=10$), we eliminated three additional items on the basis of item information. We then compared test information curves for the full 20-item version, our 13 and ten-

item versions, and three previously-reported CES-D SFs. All SFs were most useful in distinguishing among persons with depression ($\theta < 0$). Maximum information for the full 20-item version was 20.9 ($\theta = -1.8$). Maximum information for the 13 and ten-item versions occurred at $\theta = -1.8$ and was 19.1 and 17.5, respectively. The IRT-derived SFs reduced test length by 35% and 50%, respectively, while only losing 9% and 16% in measurement precision. In contrast, maximum information for previously-reported SFs was 10.1, 11.5, and 11.8, resulting in a loss of measurement precision of 44% to 52%. Using IRT methods, we derived two CES-D SFs that retain maximum information and measurement precision for use in general populations.

Abstract 138

A QUALITATIVE STUDY OF DRIVERS OF TREATMENT PREFERENCE AMONG INDIVIDUALS WITH TYPE 2 DIABETES

Colleen A. McHorney, PhD, Regenstrief Institute, Indianapolis, IN, Clarise P. Hayes, PhD, Global Economic Affairs, Eli Lilly & Company, Indianapolis, IN, Lee Bowman, PhD, Global Economic Affairs, Eli Lilly and Company, Indianapolis, IN, Jennifer Myers, MSW, Health Services Research, Roudebush VAMC, Indianapolis, IN

OBJECTIVE: Patient preferences are crucial in treatment decision-making when several equally efficacious alternative treatments are available. The objective of this study was to investigate the principal drivers of treatment preference among individuals with type 2 diabetes. **METHODS:** We conducted 11 focus groups with 84 adults with type 2 diabetes supplemented with treatment preference driver checklists. **RESULTS:** The first 5 focus groups yielded 10 drivers of treatment preference. The second 6 focus groups ranked the importance of the 10 drivers among 100 points. The principal driver of treatment preference was medication effectiveness with an average score of 36.2 out of 100. The next two highly-rated drivers were treatment flexibility and physician recommendation (9.5 and 9.4, respectively), followed by quality of life impacts and correct dosing (7.5 each), financial costs (7.3), treatment convenience (6.4), physical side effects (6.3), emotional side effects (6.0), and treatment tolerability (3.8). A full 62% of participants chose 5 or more drivers, while only 12 % chose one or two drivers. We then asked participants to assume medication effectiveness was perfect and to reallocate the 100 points among the remaining 9 drivers. In this round, the principal driver of treatment preference was physical side effects (17.4). The next most highly-rated drivers were financial costs and physician recommendation (12.9 and 12.2, respectively), followed by correct dosing (11.1), treatment flexibility (10.9), quality of life impacts (10.1), treatment convenience (9.4), emotional side effects (8.5), and treatment tolerability (7.3). Only 8% of participants chose one or two drivers, while 38% chose eight or more drivers. **CONCLUSIONS:** Great variability exists in the drivers of treatment preference among individuals with type 2 diabetes. Group averages mask tremendous inter-individual variability in the importance of drivers and their relative rank. These findings underscore the need for continued methodological work on the concept of treatment preference.

Abstract 139

DIFFERENTIAL ITEM FUNCTIONING IN FUNCTIONAL STATUS ITEMS

Colleen A. McHorney, PhD, Patrick O. Monahan, PhD, Timothy E. Stump, MA, Anthony J. Perkins, MS, School of Medicine, Indiana University, Indianapolis, IN

This poster reports on differential item functioning (DIF) in the 23 functional status items included in the Supplement on Aging (SOA). We also determined the impact of DIF on the total score. We used three methods to assess DIF: the Mantel-Haenszel procedure, logistic regression, and IRT-based DIF detection. Groups used for DIF analyses consisted of: females vs. males, aged 65-74, 75-84, or 85+ vs. those aged 55-64, African-Americans vs. Whites vs. Hispanics, those with 0-8, 9-11, or 12 years of education vs. those with 13+ years of education, and in poverty or not. We used the two-stage purification approach for DIF detection. In addition to statistical significance, we also characterized the magnitude of DIF by an effect size measure (definite vs. marginal in effect size). Definite gender DIF was identified in two items and marginal DIF in an additional six items. No items exhibited DIF in the comparison of age 55-64 vs. 65-74. Two items displayed definite DIF in the 55-64 vs. 75-84 comparison, but 13 items had definite DIF in the 55-64 vs. 85+ comparisons. Two items showed definite DIF when comparing African-Americans vs. Whites. Only marginal DIF was displayed on three items when comparing Whites vs. Hispanics. Two items had definite DIF across the education groups. Three items displayed marginal DIF between the poverty groups. After

creating total test scores with and without definite DIF items, we found that the removal of DIF items only slightly attenuated group differences. Functional status items are susceptible to item-level DIF, although test-level bias did not emerge in the SOA sample. Functional status tools that lack measurement equivalence across diverse subpopulations could result in flawed research and erroneous policy and clinical decisions.

Abstract 140

PSYCHOLOGICAL GENERAL WELL-BEING AS A PATIENT-REPORTED OUTCOME FOR THE EVALUATION OF DIABETES DRUG THERAPY

Risa P. Hayes, PhD, Lee Bowman, PhD, Global Health Outcomes Research, Eli Lilly & Company, Indianapolis, IN

Psychological general well-being is an important aspect of the quality of life of individuals with type 2 diabetes. The objective of this study was to determine the value of assessing psychological general well-being in anti-diabetes drug therapy evaluations. We administered the Psychological General Well-Being Schedule (PGWB) to 111 patients with type 2 diabetes (mean age = 55.8 years, 62% male, baseline A1c mean = 8.2%) participating in a Phase II randomized placebo-controlled trial of oral anti-diabetes treatment. The PGWB consists of 22 items divided into 6 subscales: anxiety, depressed mood, positive well-being, self-control, general health, and vitality. Pearson product moment correlation coefficients were calculated between PGWB scales and A1c (baseline and end of study) and between changes in PGWB scales and changes in A1c. At baseline, lower A1c was associated with less anxiety, less depressed mood, positive well-being, and better overall psychological general well-being [range: $r = -0.20$ (positive well-being) to -0.23 (depressed mood), $p < 0.05$]. At end of study, lower A1c was significantly associated with overall well-being and all domains [range: $r = -0.19$ (general health) to -0.42 (vitality), $p < 0.05$]. Changes in A1c were associated with changes in vitality ($r = 0.24$, $n = 111$, $p < 0.05$) but not other aspects of psychological general well-being. Well-being is associated with patient clinical condition at baseline and end of study as reflected in A1c, but, with the exception of vitality, it is not associated with shorter-term changes in A1c. Longer term follow-up may be needed to capture the impacts of drug therapy on symptoms and complications which are more directly associated with well-being. Therefore, data collected regarding patient psychological general well being may serve as a valuable supplement to clinical endpoints in diabetes drug therapy evaluations of longer duration than 12 weeks.

Abstract 141

BODY MASS INDEX, SELF-REPORTED COMORBIDITIES, AGE, AND GENDER AND THEIR RELATIONSHIP TO OBESITY-RELATED QUALITY OF LIFE IN FOUR COUNTRIES

Risa P. Hayes, PhD, Lee Bowman, PhD, Global Health Outcomes Research, Eli Lilly & Company, Indianapolis, IN

QOL is an important outcome in the evaluation of weight loss interventions (e.g., drug therapy). The objective of this study was to determine whether there are country-specific differences in the perception of the impact of weight on quality of life (QOL) or comfort with food. The Impact of Weight on Quality of Life Scale (IWQOL) was administered to a sample of 2144 obese individuals [Mean body mass index (BMI) =31.3, 63% ≥ 1 self-reported comorbidity, mean age = 41 years, 58% female] in Germany ($n=542$), Italy ($n=503$), United Kingdom (UK) ($n=539$), and United States ($n =560$). The IWQOL consists of 74 items forming 8 subscales: Health, Social/Interpersonal, Sexual Life, Work, Self-Esteem, Mobility, Activities of Daily Living, and Comfort with Food. For each country, Pearson correlations were calculated between demographics and IWQOL scores. To control for demographic differences in country-specific samples, differences in IWQOL scores were tested using analysis of covariance controlling for BMI, total number of self-reported comorbidities, gender, and age. Correlations between demographics and IWQOL subscales across countries followed a similar pattern, but the correlation coefficients calculated for Germany between most IWQOL subscale scores and BMI or total comorbidities were significantly ($p < 0.05$) higher. Significant ($p < .05$) country differences were found for all IWQOL scales. Pairwise comparisons showed, for example, that, compared to the other three countries, the UK sample had significantly better Esteem and Mobility scores; Germany had significantly worse Social/Interpersonal scores; and Italy had significantly less comfort with food (all $p < 0.05$). Differences in the perceived impact of weight on QOL found in this study suggest the need for more research investigating cultural differences in the perception of obesity. Such differences may impact the interpretation of patient-reported outcome scores obtained in international evaluations of obesity interventions.

Abstract 142

TOWARD MORE PATIENT-CENTERED CARE IN HEART FAILURE TREATMENT

Sabina B. Gesell, PhD, Aimee Williams, BA, Research & Development, Press Ganey Associates, South Bend, IN

Patient satisfaction is now widely recognized as a legitimate outcome measure of quality of care. The Institute of Medicine (IOM) has placed it on par with patient safety, effectiveness, timeliness, efficiency, and equity in the pursuit of delivering high quality care. The IOM has recommended that quality improvement initiatives be focused on specific clinical conditions because patients experience different episodes of care as a function of their clinical condition. One of the clinical conditions the IOM has targeted is heart failure because it is one of the most common and costly conditions resulting in hospitalization. To date no broad study has investigated the experience of heart failure treatment from the patient's perspective. This paper sets out to identify the unique preferences and needs that heart failure patients feel are not being adequately addressed during their hospital stay. Data for this study were extracted from the Press Ganey National Inpatient database. All surveys received between December 1, 2002 and November 30, 2003 and coded as DRG 127 (heart failure and shock) were included in the sample, which comprised 6531 patients from 269 hospitals. All participating hospitals mailed a standardized survey to random samples of patients within one week of discharge asking for honest feedback. Cronbach's alpha for the 49-item instrument was .98. A priority index was computed to determine where resources are best directed for the improvement of patient satisfaction in this population. While lower scoring questions are likely to be cause for concern, particular attention should be paid those lower scoring questions that are also highly correlated with overall satisfaction. The following issues resulted at the top of the priority index: Including patients in decisions about their care, responding to concerns patients voice in the hospital, showing understanding for the inconvenience caused by heart failure and hospitalization, providing spiritual and emotional support, and reducing wait times for tests and treatments.

Abstract 143

DEVELOPMENT OF A SHORT QUESTIONNAIRE TO EVALUATE SYMPTOMATOLOGY IN PATIENTS UNDERGOING TOTAL KNEE REPLACEMENT

Antonio Escobar, MD, Jose M. Quintana, MD, Amaya Bilbao,, Jose I. Güenaga,, Research Unit, Hospital de Basurto, Bilbao, Spain

Domains	Minor (n=411)	Mod. Case (n=265)	Intense (n=488)	Severe (n=295)
Pain (W)	32.9	45.5	53.2	65.4
FL (W)	38.7	49.9	59.5	71.5
BP (SF)	66.5	40.8	35.0	35.0
PF (SF)	45.1	34.1	31.1	31.1
RP (SF)	52.1	23.3	23.3	23.3
PCS (SF)	39.5	31.1	31.1	31.1

We used RAND methodology to develop appropriateness explicit criteria for total knee replacement. The final purpose was to apply the explicit criteria to interventions performed in different hospital. Among the variables included in the algorithm, one of them (symptomatology) required to survey the patient. The goal of the study was to create a short questionnaire to measure symptomatology. Methods. Four questions were selected, based on the panel definition for that variable. A survey including those questions was sent to patients on waiting list who were to undergo a knee replacement . To evaluate the construct validity of the scale factor analysis and multiple correspondence analysis (MCA) were performed. To study the convergent validity the scale was correlated with the WOMAC and the SF-36 questionnaires. Results. The factor analysis established one domain that matched that hypothesized with factor loadings >0.50. The categories of the questions included showed graphically a clear grouping and severity scaling, from minor to severe, on the MCA, which support the development of a score on four categories: minor, moderate, intense and severe. Cronbach alpha was 0.62. The scale clearly correlated with the WOMAC (W) and the SF-36 (SF) questionnaires. Statistically differences were found among all severity categories of the scale. Conclusion. These data give support to the validity and reliability of this scale as to be used to classify patients on severity degrees according to our panel definition.

Abstract 144

A SHORT QUESTIONNAIRE FOR THE EVALUATION OF PATIENTS UNDERGOING HIP REPLACEMENT

Jose M. Quintana, PhD, Antonio Escobar, MD, Amaia Bilbao, MSc, Research Unit, Osakidetza, Galdakao, Spain, Inmaculada Arostegui, MSc, Research Unit, UPV, Galdakao, Spain, Jesus Azkarate, Research Unit, Osakidetza, Galdakao, Spain

Based on RAND methodology a panel of experts was conveyed to developed appropriateness explicit criteria for hip replacement. The final purpose was to evaluate by explicit criteria the interventions performed in different hospital. Among the variables included in the algorithm, two of them (patient pain and functional limitation) required to survey the patient. The goal of the study was to create a short questionnaire to measure both domains. For pain, two questions were selected, and three for functional limitation, based on the panel definition for each of those variables. A survey including those questions was sent to patients on waiting list who were to undergo a hip replacement procedure. To evaluate the construct validity of both scales factor analysis and multiple correspondence analysis (MCA) were performed. To study the convergent validity both scales were correlated with the WOMAC and the SF-36 questionnaires. 734 patients answered to the questionnaires. The factor analysis clearly established two domains that matched those hypothesized (item-subscale correlations >0.76 for the functional limitation domain and >0.79 for pain). The categories of the questions included in each of those two domains showed graphically a clear grouping and severity scaling, from minor to severe, on the MCA analysis, which support the development of a score for each of them on three categories: minor, moderate and severe. Cronbach alpha was 0.47 for pain and 0.68 for functional limitation. Both scales clearly correlated with the WOMAC (W) and the SF-36 (SF) questionnaires as it can be seen in the table below (means included). Statistically significant differences were found among all severity categories for both domains, except for the Pain domain with Physical FunctionSF). These data give partial support to the validity and reliability of these two scales as to be used to classify patients on severity degrees according to our panel definition.

Abstract 145

WOMEN'S SELF-REPORTED QUALITY OF LIFE AFTER PROPHYLACTIC MASTECTOMY

Sharon J. Rolnick, PhD, Research, HealthPartners Research Foundation, Minneapolis, MN, Ann M. Geiger, PhD, Carmen N. West, MHA, Amy L. Liu, MS, Girma T. Wolde-Tsadik, PhD, Research, Kaiser Permanente Southern California, Pasadena, CA

Bilateral prophylactic mastectomy (PM) in women with increased breast cancer risk and contralateral PM in women with unilateral breast cancer dramatically reduce the risk of subsequent breast cancer. However, little is known about long-term satisfaction with the procedure and related quality of life (QOL) outcomes. Surveys were mailed to 816 women with bilateral or contralateral PM done at six health maintenance organizations between 1979 and 1999. The first survey included a \$5 cash incentive. Nonrespondents received a second mailing and a telephone reminder. The survey included validated measures of general health, body image, depression, sexuality and concerns about breast cancer as well as satisfaction with their PM decision. We received completed surveys from 661 (81%) women; 115 women with bilateral and 546 with contralateral PM. For women with bilateral PM, the mean age was 57 and the mean years since procedure was 12. For women with contralateral PM, the mean age was 61 and the mean years since procedure was 10. Over 80% of respondents indicated they are very satisfied with their PM, yet half remain very concerned about their breast cancer risk, with no differences between the two types of PM. Among

women with bilateral PM, 65 (61.3%) reported being satisfied with their QOL, versus 391 (76.2%) of women with contralateral PM (chi square $p=0.03$). Younger women were less satisfied with their procedure and QOL, and had more concern about their breast cancer risk. Results did not vary by time since procedure. Satisfaction and QOL after bilateral and contralateral PM is high but not universal. Surprisingly, many women remain concerned about breast cancer despite undergoing a procedure that dramatically reduces their risk. These results may be useful for women considering PM and their clinicians, whose discussions should include clear information on risk reduction.

Abstract 146

PRELIMINARY EVIDENCE OF RELATIONSHIP BETWEEN GENETIC MARKERS AND ONCOLOGY PATIENT QUALITY OF LIFE (QOL)

Jeff A. Sloan, PhD, Biostatistics, Mayo Clinic, Rochester, MN, Howard McLeod, PharmD, Internal Medicine, Washington University, St. Louis, MO, Daniel J. Sargent, PhD, Xinghua Zhao, MS, Biostatistics, Mayo Clinic, Rochester, MN, Charles S. Fuchs, MD, Oncology, Dana Farber Cancer Center Institute, Rochester, MN, Ramesh K. Ramanathan, MD, Medical Oncology, University of Pittsburgh Cancer Institute, Rochester, MN, Stephen K. Williamson, MD, Hematology and Oncology, University of Kansas Medical Center, Rochester, MN, Brian Findlay, MD, National Cancer Institute of Canada, Rochester, MN, Roscoe F. Morton, MD, Clinical Oncology Program, Iowa Oncology Research Association Community, Rochester, MN, Richard M. Goldberg, MD., Hematology/Oncology, University of North Carolina at Chapel Hill, Rochester, MN

PURPOSE. Is there a relationship between the state of cancer patients' genetic structure and their quality of life? Genetic variants have been associated with patient toxicity or clinical outcome. There has been little work done exploring the relationship between genetic variants and patient QOL. We pose the following research question: is there any association between genetic markers in folate genes with oncology quality of life outcome? **METHODS.** 494 patients on a GI Intergroup phase III trial of metastatic colorectal cancer to investigate the competitive efficacy of combinations of 5-fluorouracil, irinotecan, and oxaliplatin provided their genomic DNA samples and completed QOL forms (Symptom Distress Scale (SDS), and visual analog scales). Three folate candidate genes were evaluated (DPYD, MTHFR, and TYMS). Two-sample procedures examined the differences in QOL between patients with genetic variants and wild-type patients. Differences of at least 10 points on a 0-100 point range were considered clinically significant and p -values less than 0.05 were considered statistically significant. **RESULTS.** Genetic variants in DPYD*5 and TYMS TSER produced patients with differences in QOL. Statistically and clinically significant differences (10.3 points; p -value 0.008) on fatigue scores were found for DPYD*5 genotype. The A/A allele of DPYD*5 had lower fatigue score ($p=0.008$). TYMS TSER marker differed on overall SDS, fatigue and on outlook score ($p=0.007$, 0.02, 0.007 respectively). No difference was found on MTHFR genes ($p>0.05$). **CONCLUSIONS.** Folate homostasis is important for cellular functions. The associations between DPYD or TYMS and QOL are encouraging. There is sufficient evidence to suggest that there may indeed be a link between genetic structure and QOL. This hypothesis-generating study involved only patients with colorectal cancer, looked at only a few markers and specifically-targeted QOL endpoints, and yet uncovered evidence of potentially strong and specific relationships. In particular, the relationship between fatigue and genetic structure is worthy of further study.

Abstract 147

PARENT REPORTED OUTCOMES FOR THE TREATMENT OF ACUTE OTITIS MEDIA WITH EITHER CEFDINIR OR AMOXICILLIN/CLAUBLANATE ORAL SUSPENSIONS

Mary A. Cifaldi, PhD, Outcomes Research, Maria Paris, MD, Karen Devcich, MS, Stanley Bukofzer, MD, Clinical Research, Abbott Laboratories, Abbott Park, IL

Introduction: The objective of this study was to compare parent reported outcomes for children taking either cefdinir or amoxicillin/clavulanate (A/C) oral suspensions for the treatment of acute otitis media. The outcomes assessed were satisfaction, tolerability, compliance and work/daycare missed. **Methods:** In a phase IV, investigator blinded, parallel-group, randomized, multi-center study designed to compare safety and efficacy of cefdinir oral suspension (14mg/kg/day, divided every 12 hours for five days) to A/C 45mg/kg/day amoxicillin base, divided every 12 hours for ten days). Parents or legally authorized representatives of patients were asked to complete the Otitis Parent Questionnaire (OPQ) 12-14 days after the first dose of treatment. **Results:** Parents reported statistically significantly better ease of use ($p=0.009$) and taste ($p<0.0001$) in the cefdinir vs the A/C group. Parents reported

their children were significantly more likely to vomit in the A/C group than in the cefdinir group, (16% vs 7%; $p=0.016$) and the vomiting bothered them more in the A/C group ($p=0.032$). Parents reported their children were more likely to take all of their medication in the cefdinir group (68%) than in the A/C group (53%), ($p= 0.005$). There were no statistically significant differences in work/daycare missed. Conclusions: Based on parents' assessment of the OPQ, cefdinir was easier to administer and better tasting. In addition, children experienced less vomiting and were more compliant than those children who received A/C.

Abstract 148

FACTORS RELATED TO LOW UTILITY OF EUROQOL EQ-5D AMONG RURAL INHABITANTS OF KYOTO, JAPAN

Yoshiyuki Watanabe, MD, PhD, Mari Fujita, Junko Nagura, MD, Kyohei Hayashi, PhD, Kotaro Ozasa, MD, PhD, Epidemiology for Community Health, Kyoto Prefectural University of Medicine, Kyoto, Japan, Chisato Hamashima, MD, PhD, Cancer Information and Epidemiology Division, National Cancer Research Institute, Tokyo, Japan

Objective: We have conducted a case-control study on the utility measured by the EuroQOL EQ-5D among rural inhabitants, to reveal factors related to low utility. **Methods:** All the inhabitants aged 30 years or older of a rural area of Kyoto were asked to answer the Japanese clinical version of the EQ-5D, including past medical history, habitual medicine use, social conditions, lifestyle habits, etc., using a self-administered questionnaire, from June 1999 through February 2000. The inhabitants with a utility calculated by the Japanese version tariff < 0.700 (Case 1) and those with a utility < 1.00 and > 0.700 (Case 2) were compared with those with a utility = 1.000 (Control) on past medical history (diabetes, hypertension, stroke, myocardial infarction, angina, etc.), habitual medicine use (painkiller use, supplement drug use, etc.), social conditions (marital status, school career, etc), and lifestyle factors (smoking, drinking, physical exercise). **Results:** A total of 4900 (70.0%) out of the 7000 rural residents responded. Among the 4407 residents who answered all five questions of the EQ-5D, Case 1, Case 2, and Control consisted of 651, 1056, and 2700 people, respectively. The odds ratios (95% confidence intervals [CI]) of stroke history for male Case 1, male Case 2, female Case 1, and female Case 2 were 4.11 (2.20 - 7.68), 2.65 (1.43 - 4.92), 3.95 (1.86 - 8.38), and 2.34 (1.13 - 4.88), respectively. The odds ratios (95% CI) of habitual painkiller use for male case 1, male case 2, female case 1, and female case 2 were 3.66 (2.57 - 5.23), 2.52 (1.94 - 3.28), 2.58 (1.92 - 3.47), and 1.76 (1.43 - 2.16), respectively. **Conclusions:** Stroke history and habitual painkiller use were significantly related to decreased utility.

Abstract 149

MAKING A PROMISE: ETHICAL IMPLICATIONS FOR QUALITY OF LIFE

Constance L. Milton, PhD, RN, Nursing, Olivet Nazarene University, Bourbonnais, IL

The purpose of this qualitative nursing research study was to explore the meaning of making a promise, a mode of living health and quality of life conceptualized from the perspective of Parse's human becoming theory. Parse's phenomenological-hermeneutic research method was used to generate the structure of the lived experience. Through dialogical engagement with the researcher, nine women and one man described the experience in diverse contexts of their choosing. Through the process of extraction-synthesis, the researcher identified three core concepts common to all ten participants' descriptions. These were linked to form the structure of making a promise as engaging with the cherished in a solemn allegiance, while persisting with fidelity amid the assuredness-unassuredness of remorse. Through heuristic interpretation, this structure was woven with the human becoming school of thought, thus adding to nursing knowledge and knowledge about quality of life for those who make solemn commitments in the making and keeping of such promises as living wills and durable powers of attorney for healthcare. Possible ethical implications for being with persons who make such vital commitments will be explored in this presentation.

Abstract 150

A STUDY OF URINARY INCONTINENCE: ITS PREVALANCE AND ITS EFFECT ON THE QUALITY OF LIFE

Hyojeong Song, PhD, Nursing, School of Medicine, Cheju National University, Jeju-shi, South Korea, myungja Kim, PhD, College of Nursing, The Catholic University of Korea, Seoul, South Korea

Purpose : This study was designed to investigate the prevalence of female urinary incontinence, and factors affecting the quality of life. It was also to provide the basic data for the primary urinary incontinence management program. **Materials and Methods :** The Study was performed from May 7 to July 18, 2000. The subjects of the study consisted of 2183 women, aged over 30 up to 89, residing in Seoul, Kyongki Kangwon, Chungchong, YongNam, HoNam, Cheju provinces. Those who understood and responded to the questionnaires were selected from church meetings, teaching facilities, corporations, public offices, the voluntary organizations, and institutes for old people. The data were analyzed by Student t-test, ANOVA, and stepwise regression with using SAS program. **Results :** The overall reported prevalence of urinary incontinence was 55.7% with urinary leakage more than once a month for the past one year. In regarding age, the highest prevalence was 71.0 % in the fifties. The stress urinary incontinence was 60.8 %, the mixed urinary incontinence 38.2 %, the urge urinary incontinence 1.0 %. The urinary incontinence related quality of life indicated significant differences by age($F= 25.69$, $p= 0.0001$), the educational status($F= 27.86$, $p= 0.0001$), the economic status($F= 8.14$, $p= 0.0001$), body mass index($F= 1.27$, $p= 0.0001$), the urinary incontinence pattern($F= 115.19$, $p= 0.0001$), the urinary incontinence severity($F= 105.00$, $p= 0.0001$), the perceived urinary symptom($F= 90.87$, $p= 0.0001$). The urinary incontinence related quality of life was significantly predicted by lower urinary tract symptoms(25.3%), the number of delivery, the urinary incontinence severity, the family support and the self efficacy in sequence. These variables explained 38.0% of the variance of urinary incontinence related quality of life. **Conclusion :** The above findings indicated that the proper nursing intervention for lower urinary tract symptoms, the family support and the self efficacy was recommended to promote the quality of life for the persons with urinary incontinence

Abstract 151

USING PARTIAL LEAST SQUARES (PLS) ANALYSIS TO SHORTEN A PREDICTIVE SYMPTOM CHECKLIST FOR UPPER GASTRO-INTESTINAL (UGI) DISORDERS

Benoit Arnould, MSc, Individual PRO Tools, Mapi Values, Lyon, France, Dominique Dubois, PhD, PGSM Health Economics, Johnson&Johnson Pharmaceutical Services, Turnhoutseweg 30, Belgium, Antoine Regnault, MSc, Laboratoire d'analyse des systèmes de santé, Université Claude Bernard Lyon 1, Lyon, France

Introduction: The maturity of patient-reported outcomes (PRO) methodology triggers an increasing demand by general practitioners (GPs) for easy-to-interpret and validated measures of the patients own perception of symptoms. Symptom checklists may be particularly useful tools for the diagnosis of UGI disorders, where the interpretation of various symptoms is essential. **Objective:** To develop, score and validate a short symptom checklist that predicts the GPs diagnosis. **Patients and Methods:** 352 patients included by GPs in USA, Australia, Italy and the NL, were diagnosed as GERD (n=214), Ulcer-like Dyspepsia (n= 76), or Dysmotility-like Dyspepsia (n= 62). Patients had to complete a checklist of 148 items covering 37 symptoms rated in frequency, intensity, and bothersomeness on 5-points likert scales. Three analytical methods were used to select the items and to build predictive models: linear discriminant analysis, Partial Least Squares (PLS) discriminant analysis and polytomous logistic regression. The capacity of the models to discriminate the three diagnosis groups was measured by the volume under the ROC surface. A model with a value >0.16 has better discriminant power than random. **Results:** The models included 7 to 24 items, covering 5 to 10 symptoms. Only one symptom was selected in all models. All the predictive models had good discriminant validity (0.40 to 0.67). The concordance with GPs ranged from 47% to 63%. The simplification of the models to hand-calculated scores did not impair their predictive validity. The robustness of the models was good for the PLS method, and moderate for the other models. **Discussion:** This study shows that an easy to score symptom checklist completed by the patient has very encouraging performances to predict the GPs diagnosis. The choice of the method, as well as the contribution of clinicians, are essential to ensure the validity and robustness of the predictive model. The next step will be to assess these performances against an objective diagnostic gold standard.

Abstract 152

PROGNOSIS AND QUALITY OF LIFE IN PATIENTS WITH ACUTE LOW BACK PAIN. INSIGHTS FROM A COMPREHENSIVE INCEPTION COHORT STUDY

Joel Coste, MD, PhD, Biostatistique, Hôpital Cochin, Paris, France, Gérard Lefrançois, MD, Affaires Médicales, Laboratoire Chiesi, Courbevoie, France, Jacques Pouchot, MD, PhD, Service de Médecine Interne, Hôpital Louis Mourier, Colombes, France, Francis Guillemin, MD, PhD, Ecole de Santé Publique, Faculté de Médecine, Vandoeuvre les Nancy, France

There is increasing evidence that a biopsychosocial approach is more appropriate than biomechanical considerations for the analysis of low back pain (LBP). We conducted an inception cohort study of LBP patients to investigate the respective contribution of biological and psychosocial factors, and especially initial health-related quality of life (HRQOL), in the natural history of acute LBP and to evaluate the impact of this condition on HRQOL. We followed for three months 113 patients consulting for nonspecific acute LBP of less than 72 hours duration at inclusion and treated with paracetamol (acetaminophen). Endpoints included pain, disability assessed by the Roland disability questionnaire, and HRQOL assessed by the 36-item Short-Form General Health Survey (SF-36). 73% of patients recovered within two weeks and 5 % of patients developed chronic LBP. Prior low back surgery, higher initial disability questionnaire score, lower SF-36 general health score, and compensation status were independently associated with delayed recovery. The impact of the acute LBP episode on HRQOL was brief and moderate, except for patients with comorbidity and psychiatric disorders, and those of foreign origin, unemployed, or with job dissatisfaction. The impact of compensation status, sick leave and bedrest was more profound and lasting. This study highlights the large contribution of psychosocial factors, and especially initial HRQOL to the prognosis of LBP. It also suggests a vicious circle, in which LBP impairs HRQOL, through societal recognition and stigmatizing medical care, which in turn favors the condition becoming chronic. These findings have implications for future research in the management of patients with LBP.

Abstract 153

THAI PATIENTS' EXPERIENCE OF THE BURDEN OF ASTHMA: QUALITATIVE STUDY

Preeyanut Sirimai, Mpharm, Pharmacy, Prachuapkhirikhan Hospital, Muang District, Prachuapkhirikhan, Thailand, Chulaporn Limwattananon, PhD, Supon Limwattananon, Pharmacy, Khon Kaen University, Khon Kaen, Thailand, Viroj Tangcharoensatien, PhD, International Health Policy Program, Bangkok, Thailand, Watchara Boonsawat, PhD, Medicine, Khon Kean University, Khon Kaen, Thailand

Objectives: To describe the burden of asthma among Thai patients. Design: Qualitative study with semi-structured in-depth interviews. Setting: A secondary hospital in the upper south of Thailand. Participants: 52 asthmatic patients aged 18-60 years experiencing in emergency department visits in 2003. Results: The burden of asthma affected Thai patients on mainly 4 categories relating to daily life including physical and emotional impact, employment, financial, and partners relationships. Suffering from asthma symptoms was stated in most participants with poorly controlled asthma. Emotions of fear and panic appeared from asthma attack in all patients. Suicidal idea from disease and financial problems took place in some participants. In persistent asthmatic patients, feelings of exhaustion and tiredness frequently happened and caused of omitting therapy in some participants. Financial problems occurred from losing or changing jobs or losing income affecting patients who were chief of the family. Partner relationships could end up with divorce due to financial problems. The cost of asthma medication was not an issue in most patients since universal coverage was implemented at a nationwide scale. Conclusions: Our current understanding of the patients' experience of the burden of asthma contributes to improving the management and control of asthma and enabling patients to live full, active and productive lives.

Abstract 154

EVALUATING THE EFFECTS OF ONCOLOGICAL REHABILITATION INTERVENTIONS USING EFFICACY INDICES CALCULATED FROM PATIENT-REPORTED OUTCOMES

Michael T. Moser, Dipl.Psych., Joachim Weis, Prof, Hans Helge Bartsch, Prof, Institute for Rehabilitation Reserach, Tumor Biology Center, Freiburg, Germany

Interdisciplinary rehabilitation programs for cancer patients were recognized as important and necessary interventions because positive effects on QoL, relapse and survival were reported. The effects of these programs has rarely been explored under controlled conditions. Interdisciplinary rehabilitation programs take into account all medical and psycho-social aspects of oncological aftercare. Patient-reported QoL assessment is the best method to evaluate these programs, because QoL instruments cover a broad range of outcome domains. Therapeutic goals in oncological rehab programs are very heterogeneous. Depending which therapeutic goals are indicated, the concordance of individual goals and outcome domains could vary and treatment effects could be underestimated in studies with multiple QoL endpoints. Our findings are based on a group of patients participating in a prospective controlled multi-center trial for evaluation of inpatient rehab programs in Germany. 351 women with non-metastatic

breast cancer were included in the rehabilitation group and 103 patients in the non-treatment group. QoL was assessed with the EORTC-QLQ-C30 and breast module at admission, discharge (resp. week four) and after one year. MANCOVA analyses showed significant beneficial effects for the treatment group in central QoL-domains ($p < .03$) over time. To demonstrate the clinical significance of these findings, two methods for computing individual change rates were employed: 1. reliability-based approaches (reliable change statistics) and 2. normative approaches (change rates indicating an individual transition into another health state using normative data). Change rates between the rehabilitation group and the controls were compared and used to calculate established measures for treatment efficacy (risk indices or NNT). Comparative analyses between these two methodological approaches revealed convergent estimations for the benefit of the intervention programs.

Abstract 155

TESTS OF SCALING ASSUMPTIONS, CONSTRUCT VALIDITY AND RELIABILITY OF THE CHINESE CHILD HEALTH QUESTIONNAIRE, PARENT FORM (CHQ-PF50) AND CHILD FORM (CHQ-CF87)

Yuen-Yee J. Ng, MSc, Physiotherapy Department, Grantham Hospital, Aberdeen, China, Jeanne M. Landgraf, MA, Health Act, Health Act, Boston, MA, Shui Wah Chiu, Consultant, Cardio-thoracic Surgery, Grantham Hospital, Aberdeen, China, N Cheng,, Yiu Fai Cheung,, Paediatrics, The University of Hong Kong, Hong Kong, China

Validated instruments for assessment of quality of life in Chinese children are as yet unavailable. We determined the construct validity and reliability of the translated Chinese versions of the Child Health Questionnaires (Landgraf, Abetz & Ware, 1999) designed for completion by parents (CHQ-PF50) and children (CHQ-CF87). The Chinese versions were developed through iterative forward and backward translation processes by independent parties. The feasibility, as rated by degree of difficulty using a 4-point scale, and time for completion were evaluated for the Chinese CHQ-PF50 and CHQ-CF87 in 15 and 11 subjects, respectively. To assess the construct validity and reliability, 1143 parents of healthy children and 823 school children were invited to complete the Chinese CHQ-PF50 and CHQ-CF87, respectively. The results showed that both the Chinese CHQ-PF50 (mean rating 1.66) and CHQ-CF87 (mean rating 1.33) were easy to complete, with completion times of 14.23 ± 5.23 minutes and 13.82 ± 3.52 minutes, respectively. Psychometric analysis on item convergent validity and discriminant validity showed perfect or near perfect (>99%) rates of success for all ten scales in the CF87 and >94% for all but one scale in the PF50. The exception was the general health scale (86%). Minimal floor effects were observed for both questionnaires. However, substantial ceiling effects were observed for the five scales in both questionnaires (physical functioning, role-emotional, behavioral, role-physical, bodily pain and family activities). The median alpha coefficient of reliability for CF87 was 0.90 (range 0.85 to 0.94). The median alpha coefficient for PF50 was 0.80 (range 0.44 to 0.88), with the mental health scale falling just below the minimum criterion for group level analysis (0.68) and the general health scale being the lowest (0.44). These findings suggest that the Chinese translations of CHQ-PF50 and CHQ-CF87 are robust and sufficient. Additional work with regard to ceiling effects is required to assess the performance of the measures in condition groups.

Abstract 156

AGE AND GENDER DIFFERENCES ON HEALTH-RELATED QUALITY OF LIFE (SF-36) FOR PEOPLE AFTER CARDIAC SURGERY/INTERVENTION

Yuen-Yee J. Ng, MSc, Physiotherapy Department, WH Chow, Cardiac Medicine, SW Chiu, Cardio-thoracic Surgery, SC Leung, HCE, Grantham Hospital, Aberdeen, China, SF Tam, Rehabilitation Sciences, The Hong Kong Polytechnic University, Kowloon, China

Quality of life represents the multi-factorial effect of a disease and/or an intervention and reflects the therapy consequences as perceived by a patient. The purpose of this study was to investigate the age and gender effect on life quality among female and male, adults and elderly 4-6 weeks after cardiac surgery/intervention. A convenient sampling of totally 262 participants, 194 males and 68 females, with age ranged from 21 to 84 (mean=57.31, S.D.=11.7) were invited to join the study when they entered an out-patient cardiac rehabilitation program. They were requested to self-administer the SF-36 questionnaire. Their physical endurance was measured by 6-minute walking test. The results were analyzed by using a 2 (median-split of age) X 2 (gender) ANOVA factorial analysis with education level and 6-minute walking test performance as covariates. The results showed that there were significant age x gender effects on physical function ($F=3.60$, $p < 0.05$), social function ($F=12.02$, $p < 0.01$) and role emotion ($F=3.89$,

$p < 0.05$). Gender differences were found within the elderly group on social function ($F = 6.59$, $p < 0.01$) and role emotion ($F = 8.06$, $p < 0.01$) but not within the adult group. Age and gender differences were also evidenced on 6-minute walking test. Physical endurance of elderly was significantly lower than adults ($t = 6.116$, $p < 0.001$). Women's physical endurance was significantly lower than men's ($t = 5.797$, $p < 0.001$). The elderly with cardiac disease perceived better life quality than adults with less pain ($F = 7.52$, $p < 0.01$), better general and mental health ($F = 8.15$ and 7.10 , both at $p < 0.01$), better vitality ($F = 12.62$, $p < 0.01$) and less limitation caused by emotion ($F = 8.84$, $p < 0.01$) though they had poorer physical endurance. Elderly women had the highest role emotion but lowest physical and social functions. When comparing to the norm, obvious life quality depletion was found mostly on physical function, role physical, bodily pain and social function (all below 25th percentile).

Abstract 157

U.S. VALUATION OF THE EQ-5D HEALTH STATES: DEVELOPMENT AND TESTING OF THE D1 VALUATION MODEL

James W. Shaw, PharmD, Pharmaceutical Sciences, University of Arizona, Tucson, AZ, Jeffrey A. Johnson, PhD, Institute of Health Economics, University of Alberta, Edmonton, AB, Canada, Stephen J. Coons, PhD, Pharmaceutical Sciences, University of Arizona, Tucson, AZ

The EQ-5D is a brief, multi-attribute, preference-based health status measure. This paper describes efforts to develop a statistical model for estimating U.S. population-based preference weights for the EQ-5D. A multistage probability sample was selected from the adult U.S. population. Each respondent was randomly assigned 15 health states from a subset of 45 of the 243 possible states described by the EQ-5D. Respondents valued 13 of these states using the time trade-off (TTO) method. The TTO valuations were linearly transformed to lie on the interval $[-1, 1]$. Numerous models were investigated to account for interaction effects due to having problems in multiple dimensions. Several alternative specifications (e.g., pooled least squares, random effects) were also considered. A modified split-sample approach was used to evaluate the predictive accuracy of the models. All statistical analyses took into account the clustering and disproportionate selection probabilities inherent in our sampling design.

As found in previous studies, models including a dummy variable for extreme disability in any EQ-5D dimension provided a reasonable fit for the data. However, the best model proved to be one based on a conceptual notion of the effect of movements away from perfect health. This model, which we have named D1, included ordinal terms to capture the effect of departures from perfect health as well as interaction effects due to increasing health problems. A random effects specification of the D1 model yielded a good fit for the observed TTO data, with an overall R^2 of 0.38, a mean absolute error of 0.02, and a correlation between mean observed and predicted valuations of 0.99.

The D1 model predicts the values for observed health states with a high degree of accuracy. This model's estimates will provide a set of EQ-5D preference weights specifically developed for use in the U.S. population.

Abstract 158

CLINICALLY ASSOCIATED DIFFERENCES IN SELF-REPORTED VISUAL FUNCTIONING BY LATINOS

Denise R. Globe, PhD, Joanne Wu, MS, Pharmaceutical Economics and Policy, Rohit Varma, MD, Keck School of Medicine, University of Southern California, Los Angeles, CA

Purpose: To assess differences in NEI-VFQ-25 scores that are associated with clinically relevant differences in visual acuity. **Methods:** Data from the Los Angeles Latino Eye Study (LALES), a population-based, cross-sectional prevalence study of eye disease in Latinos aged 40 and older, residing in Los Angeles, California were used. All participants completed a standardized interview, including the 25-Item NEI VFQ-25 to measure visual functioning, and a detailed eye examination. Visual acuity was measured and scored as the total number of letters read correctly and converted to a log of the minimum angle of resolution (logMAR) scores. To estimate a clinically relevant NEI VFQ-25 difference score, we used a 0.2 logMAR difference (equivalent to a clinically relevant two line difference on the visual acuity chart) in visual acuity as the anchor. The parameter estimates of the regression model for each subscale score was used to determine the number of points for each subscale score associated with a constant difference of two lines of acuity. **Results:** A constant two-line difference in visual acuity was associated with a two (Ocular Pain) to nine point (Driving Difficulties) difference in subscale scores. **Conclusion:** These findings suggest that the magnitude of differences in visual functioning associated with visual impairment varies by the content

area of the subscale. Two lines of difference have a greater impact on driving and distance vision compared to ocular pain or color vision. Although this study provides an estimate of what one might expect with changes over time, these changes need to be assessed in longitudinal studies.

Abstract 159

DEVELOPMENT OF A PREFERENCE WEIGHT MULTIATTRIBUTE HEALTH OUTCOME MEASURE FOR PATIENTS WITH RHEUMATOID ARTHRITIS

Chiun-Fang Chiou, PhD, Health Economics, Zynx Health, a Cerner Company, Beverly Hills, CA, Maria E. Suarez-Almazor, MD, Health Sciences Research, Baylor College of Medicine, Houston, TX, Chih-Hung Chang, PhD, Feinberg School of Medicine, Northwestern University, Northbrook, IL, Michael Weisman, MD, Medicine, Cedars-Sinai Medical Center, Los Angeles, CA

Subscale	# Points Difference (se)
Color Vision	8.90 (0.32)
Vision Related Dependence	8.20 (0.35)
Driving Difficulties	7.34 (0.28)
Distance Vision	7.34 (0.28)
General Vision	4.50 (0.26)
Vision Related Mental Health	6.98 (0.34)
Near Vision	4.43 (0.33)
Ocular Pain	4.65 (0.33)
Peripheral Vision	3.22 (0.33)
Vision Related Role Function	5.34 (0.34)
Vision Related Social Function	4.76 (0.23)
Composite	5.48 (0.23)

Quality-Adjusted Life Years (QALYs) are usually the sum of the products of the duration and interval-scaled preference weight of each health state. The durations can be measured in clinical trials, but it is usually difficult to find preference weights for measuring QALYs. We developed a multiattribute measure to assess, and assign preference weights to RA patients' health states. A steering committee selected 28 items from previous HRQoL instruments, patient focus groups, and experts' input from a recent study developing a new RA HRQoL instrument. The 28-item questionnaire was mailed to 748 patients recruited from Southern California along with visual analogue scale questions on preference for different health states, a time-trade-off question on preference for patient general health, the SF-36 questionnaire, and others. Factor analysis, an item-response theory-based model, and an internal consistency test were used to identify attributes and evaluate items retained for the new measure. We constructed 4 multiattribute preference weight functions (MAPWF) for assigning preference weights for patient-reported health states, and used R^2 and Cp values to select the best MAPWF for the new measure. A total of 487 patients returned the survey. From results of factor analysis, the IRT-based model, and the internal consistency test, 24 items were selected to form the new measure. They assess RA impact on 6 health attributes: physical function (4 items), RA symptom distress (6 items), social interaction (4 items), therapy-related symptom distress (3 items), dexterity (3 items), and emotion (4 items). The MAPWF with the dependent variable "predicted HUI-derived preference weight" was selected as the final algorithm. The new measure helps users assess RA patients' health states on 6 attributes. It is more comprehensive than most conventional RA measures. Its preference weight algorithm lets users assign preference weights to RA health states, allowing comparison of effectiveness of medical interventions for RA patients.

Abstract 160

VALIDATION OF A PREFERENCE WEIGHT MULTIATTRIBUTE HEALTH OUTCOME MEASURE FOR PATIENTS WITH RHEUMATOID ARTHRITIS

Chiun-Fang Chiou, PhD, Health Economics, Zynx Health, a Cerner Company, Beverly Hills, CA, Maria E. Suarez-Almazor, MD, Health Services Research, Baylor College of Medicine, Houston, TX, Chih-Hung Chang, PhD, Feinberg School of Medicine, Northwestern University, Northbrook, IL, Michael Weisman, MD, Medicine, Cedars-Sinai Medical Center, Los Angeles, CA

A new measure—24 items assessing RA impact on 6 health attributes including physical function, RA symptom distress, social interaction, therapy-related symptom distress, dexterity, and emotion—has been developed to allow effectiveness comparison in Quality-Adjusted Life Years (QALYs) among various medical interventions. To test the new measure's convergent validity, discriminant validity, and internal consistency. Data were obtained from a survey conducted in Southern California, USA, in the summer of 2003. Response rate was 65% (487/748): 79% female; mean age 59 (± 14) years. Pearson correlation coefficient (PCC) was used to test the convergent validity of the new measure with the MOS SF-36, the HAQ, pain and tender joint count, and swollen joint count. A PCC ≥ 0.4 would indicate that the new measure has reasonable convergent validity. Analysis of variances (ANOVA), and the Kruskal-Wallis test were performed to examine the know-group discriminant validity. Cronbach's alpha coefficient was calculated to assess the internal consistency of items measuring each attribute. A Cronbach's alpha coefficient ≥ 0.70 would indicate that a good internal consistency exists. The estimated preference weight correlated well with 2 widely used and reliable instruments, the HAQ and the MOS SF-36, as well as the pain and tender joint count and the swollen joint count (PCCs 0.79 to 0.42 in absolute values). Correlation with the joint counts was moderate though still significant ($P < 0.0001$). When patients were grouped by HAQ score, groups differed significantly (ANOVA and Kruskal-Wallis test [$P < 0.0001$]). Cronbach's alpha for all 6 attributes were > 0.70 . These results support the convergent validity, discriminant validity, and internal consistency of the new measure for assessing health states and corresponding preference weights for health states of RA patients. Other psychometric characteristics of the measure will be tested in future studies.

Abstract 161

CRITICAL APPRAISAL OF DEVELOPMENTAL AND PSYCHOMETRIC CHARACTERISTICS OF MULTI-MODAL DISEASE ACTIVITY INSTRUMENTS

Michael Broder, MD, MSHS, Lionel Pinto, MS, Anacleto Gano, Jr., MPH, Enkhe Badamgarav, MD, MPH, Life Sciences, Zynx Health - A Cerner Company, Beverly Hills, CA, Deborah Lubeck, PhD, Health Economics & Outcomes Research, Greg Rossi, PhD, Medical Affairs, Amgen Inc., Thousand Oaks, CA

To evaluate the possible utility of an index designed to assess the impact of anemia, we performed a systematic review of published literature to identify and review the developmental and psychometric properties of multi-modal instruments. We defined multi-modal as a scale incorporating a combination of self-reported items, physician assessments and objective measures. MEDLINE databases were searched for English-language articles published between 1966 and 2003, containing the search terms composite, multi-modal, and disease score among others. The review was conducted in three stages: title, abstract and full-text review. An expert panel verified the multi-modal format of identified instruments. Interrater agreement was assessed using the kappa statistic. Seven instruments were identified: 3 for systemic lupus (SLAM, SLEDAI, ECLAM); 2, rheumatoid arthritis (ACR, SDAI); one, Crohn's disease (CDAI); and one, Alzheimer's (ADAS). Lupus scales lacked patient-reported items, but the Crohn's disease scale was dominated by such items. The RA scales had a preponderance of laboratory measures. For scale development, the initial item selection was generally based on expert panel consensus. Listed items were further narrowed down using multivariate regression analysis, with physician's global assessment of the disease used as the dependent variable or 'gold standard'. Overall scores were calculated as either the simple or weighted (using regression parameters) sum of the selected items. The CDAI and ACR, which had good validity and reliability, were cited in more than 100 publications limited to clinical trials. Despite the heterogeneity of the items included in the formation of the multi-modal scales, a fairly consistent process of scale development and scoring was identified. While some have been widely utilized in clinical research, many have not gained acceptance in clinical practice.

Abstract 162

USING AN ONLINE SURVEY TO EVALUATE PATIENT SATISFACTION WITH ANTIHYPERTENSIVE MEDICATIONS

Kristina Chen, PharmD, MS, Chiun-Fang Chiou, PhD, Research, Zynx Health Incorporated, Beverly Hills, CA, Craig A. Plauschinat, PharmD, Health Economics & Outcomes Research, Novartis Pharmaceuticals Corp., East Hanover, NJ, UK, Feride H. Frech, MPH, RPh, Health Economics & Outcomes Research, Novartis Pharmaceuticals Corp., East Hanover, NJ, Robert Dubois, MD, PhD, Anne Harper, PhD, Research, Zynx Health Incorporated, Beverly Hills, CA

The Internet offers a new way of collecting patient-reported outcomes. Conducting surveys through mailing, phone, or in person is a less cost-effective process than using the Internet. The former methods of data collection require applying for institutional review board (IRB) approval, identifying the most adequate patient recruitment method and implementing the survey. We conducted a case study of adapting an online survey using a national patient panel to evaluate patient satisfaction with their antihypertensive medications. The survey was launched in September 2003 to gather rate of adverse effects, likelihood of continuing medication, drug switching behavior, and reasons for switching from a chronic disease panel. The survey was sent via email to almost 200,000 members of the panel self-reporting hypertension. A screening survey was used to identify patients receiving either diuretics, angiotensin II receptor blockers, ACE inhibitors, calcium channel blockers or beta-blockers, and to exclude patients on multiple therapies as well as those receiving the antihypertensive medication for less than one month. To increase generalizability of our sample, data were adjusted with weights obtained from the NHANES database. Pairwise comparisons among different drug classes were conducted with Tukey adjustment. Since patients had previously consented to participate in the panel, the IRB process was handled and maintained by the company who owns the panel. A pilot test was conducted in a sample of the panel to identify and correct problems with the survey questions. This phase, including revisions to the survey, and fielding of the final survey took approximately three weeks. Surveys were completed by 1,256 panel members (23% response rate) with a mean age of 56.8 +/- 14.2 years old, and 57.9% were female. The survey captured differences by drug class in all parameters measured. Online survey proved to be a cost-effective approach for collection patient-reported outcome data. Information obtained from this study may aid physicians in medication selection and ongoing monitoring for patients

Abstract 163

TYPE 2 DIABETES & LIFE SATISFACTION, THE RANCHO BERNARDO STUDY

Kristin M. David, MPH, Deborah L. Wingard, PhD, Donna Kritz-Silverstein, PhD, Elizabeth Barrett-Connor, MD, Family & Preventive Medicine, University of California, San Diego, La Jolla, CA

Self reported Health Related Quality of Life (QOL) has been shown to be lower in people with type 2 diabetes; while having health complications associated with diabetes (e.g., stroke and CHD) contributes to a lower reported QOL. The relation between type 2 diabetes and diabetes complications with self-reported life satisfaction has not been examined. The goal of the present study was to examine the relation of type 2 diabetes with and without complications to self-reported life satisfaction. Participants were 382 men and 520 women aged 50 to 80 from the Rancho Bernardo Cohort who attended a clinic visit between 1984-1987 where an oral glucose tolerance test was administered after a minimum 8 hour fast, and who completed a mailed questionnaire between 1992-1993 that included two standardized life satisfaction measures: the Life Satisfaction Index-Z (LSI-Z) and the Satisfaction with Life Scale (SWLS). Based on WHO 1999 criteria, participants were categorized as having type 2 diabetes if their fasting plasma glucose was > 126 mg/dl or their post-challenge plasma glucose level was > 200 mg/dl or they had been diagnosed by a physician as having diabetes. No differences were found for LSI-Z or SWLS between men and women with and without diabetes (p 's >.05), (nor were differences found between those with and without complications, p 's>.05) However, men with diabetes reported poorer overall health than men without diabetes (p =.04). Both men and women with diabetes also reported poorer health as compared to others than men (p =.03) and women without diabetes (p =.04). This study suggests that although men and women with type 2 diabetes may recognize that their overall health and health compared to others may be worse, it does not impact their life satisfaction.

Abstract 164

QUALITY OF LIFE SCORES AMONG ENGLISH AND SPANISH SPEAKING PATIENTS AT A FREE COMMUNITY HEALTH CLINIC

Erik J. Groessl, PhD, Theodore G. Ganiats, MD, Kristin M. David, MPH, Robert M. Kaplan, PhD, Family & Preventive Medicine, UCSD - Health Outcomes Assessment Program, La Jolla, CA, Ellen Beck, Hlth Outcomes Assessment Prgrm, Univ of CA-San Diego, San Diego, CA

Many people in the United States do not have health insurance and lack sufficient resources to pay for health services. This may result in lower quality of life, less usage of primary care services, and higher usage of emergency and urgent care services. Hence, it is important to understand the characteristics of this population. The purpose of this study is to describe and compare the quality of life of English and Spanish-speaking patients at a free community health clinic. The patients were 354 adults who received care and completed an initial Quality of Well-Being Self-administered (QWB-SA) scale at a free primary care clinic run by UCSD at 3 locations in the San Diego area. The participants were 58% female, 45 years old on average; 19% had at least a college degree, 43% were Hispanic, and 43% were Caucasian - Non-Hispanic. 17% of patients chose to complete the Spanish version of the QWB-SA instead of the English version. No significant differences were found ($t = -0.580$, $p = .562$) between QWB scores (or standard deviations) for those completing the English (0.595, $sd = 0.176$) versus the Spanish version (0.609, $sd = 0.149$) even though patients completing the Spanish version were more likely to be female, and less educated. In addition, patients completing the English version were more likely to report impairment on the social activity subscale of the QWB. The free clinic patients had lower quality of life scores (0.600) than adults in a previous study with similar demographics from a standard UCSD family medicine clinic (0.650). Patients completing the English vs. Spanish versions of the QWB-SA had similar quality of life scores and standard deviations. Patients completing the Spanish version were more likely to be female and less educated, which have been associated with lower QWB-SA scores. However, Spanish speaking patients report fewer social activity limitations, which may offset the other factors. Understanding the characteristics of free clinic patients may inform future research and interventions.

Abstract 165

COMPARISON OF THE QWB-SA AND THE HUI-3 IN A GERIATRIC POPULATION

Theodore G. Ganiats, MD, William J. Sieber, PhD, Family & Preventive Medicine, University of California San Diego, La Jolla, CA, James W. Mold, MD, Audrea M. Roberts, MA, Family Medicine, Oklahoma University School of Medicine, Oklahoma City, OK, Robert M. Kaplan, PhD, Family & Preventive Medicine, University of California San Diego, La Jolla, CA

There are a number of utility- or preference-based measures that are used to assess health-related quality of life (HRQOL) and calculate Quality-adjusted Life Years (QALYs). However, these measures can produce rather different results, and their properties should continue to be evaluated. The properties of the Self-Administered Quality of Well-being Scale (QWB-SA) and the Health Utilities Index Mark 3 (HUI-3) were compared at Years 1, 2, and 3 of a longitudinal study in a geriatric population. HRQOL was measured using the QWB-SA and the HUI-3 in 799 geriatric outpatients. The scales have different scoring systems and derive health state values using different methods, but both produce a single total score. The HUI-3 allows for health states less than 0 while the QWB-SA does not. Participants were 56% female, 86% Caucasian, and had a mean age of 73.4 years at Year 1. Descriptive statistics showed that the HUI-3 had consistently higher scores across the 3 time points, a wider range of scores and higher standard deviations, indicating more variability. At Year 3, 23% scored 0.90 or above on the HUI-3 while only 2% scored above 0.90 on the QWB-SA, indicating a possible ceiling effect for the HUI-3 in this elderly population. Correlations between the measures ranged from 0.505 to 0.537 ($p < 0.001$). QWB-SA scores across all participants showed a significant increase between Year 1 and 2 (0.634 to 0.686) while the HUI-3 showed no significant change between Year 1 and 2 (0.717 to 0.715). The results indicate that the QWB-SA has less variability, is more consistent over time, and is more normally distributed at Year 1 before attrition occurred. The QWB-SA may be better able to detect change because it has less variability and less constriction at the upper end of the distribution.

Abstract 166

SHORT-TERM PSYCHOSOCIAL COUNSELING IN PATIENTS WITH NEWLY DIAGNOSED PROSTATE CANCER

Natalia Sadetsky, MD, Andrew Kneier, PhD, David M. Latini, PhD, Sara J. Knight, PhD, Jeff Loi, BS, Peter R. Carroll, MD, Urology, UCSF, San Francisco, CA

Prostate cancer can have significant physical, psychological, and social impact on patients, particularly early in the treatment course. While psychological intervention during oncologic care has been shown to improve all aspects of well-being, few studies have focused on men diagnosed with prostate cancer or have targeted the period between diagnosis and treatment. We evaluated the efficacy of an innovative, brief psychosocial counseling program delivered to men before and after prostate cancer treatment. 132 men diagnosed with prostate cancer were randomly assigned to either intervention (n=69) or control (n=63) groups. Participants completed the Profile of Mood States (POMS), Index of Coping Responses (ICR), and scales measuring general and disease-specific HRQOL, at baseline (before treatment commenced) and at months 3 and 12. Both groups received standard medical care; and intervention participants received 2 sessions of counseling before treatment and one session 4 to 6 weeks after treatment completion. Intervention effects were evaluated using a mixed model analysis, adjusting for cancer treatment (radiation or surgery) and time. There were no significant differences between groups at baseline on the demographic or psychosocial measures. In the mixed model, intervention participants reported significant improvements over time on three outcomes: Vigor (POMS; $p < .05$) and Information Seeking and Affective Regulation (ICR; $p < .01$ and $p < .05$, respectively). While the brief intervention did not appear to influence HRQOL, or prostate-specific symptoms, it improved energy level and coping responses, particularly those relevant to managing information and emotions. These coping strategies may be especially important given the decision-making and treatment initiation tasks of early prostate cancer care. Supported by California Cancer Research Program grant # 97-12013.

Abstract 167

QUALITY OF LIFE IN SCHIZOPHRENIA WITH RECENT-ONSET: PATIENTS-REPORTED OUTCOMES ON GROUP FAMILY PSYCHOEDUCATION

Lucie Motlova, MD, School of Public Health, UC Berkeley, Berkeley, CA, Eva Dragomirecka, PhD., Filip Spaniel, MD, Richard Zalesky, MD, Eva Goppoldova., Pavla Selepova., Social Psychiatry, Prague Psychiatric Center, Prague, Czech Republic

The enhancement of quality of life and reduction of the social impairment resulting from schizophrenia can be achieved in the complex family psychoeducational program. The goals of the study were to identify the differences between quality of life of patients and general population, to investigate if those possible differences could be influenced by family psychoeducation and to collect the information on patient-identified benefits of psychoeducation. 39 out-patients (age mean 29.87, mean number of schizophrenic episodes 1.33) were administered the Subjective Quality of Life Analysis (SQUALA) questionnaire before the program and 3 months later. Their QOL scores before the program were compared with the QOL of 78 demographically matched controls (case-control design). The Psychoeducation Outcomes Questionnaire (POQ) was mailed to all participants one-year after the program. The patients with schizophrenia had significantly lower overall quality of life (QOL) ($p < 0.05$) compared to their healthy counterparts. The patients improved significantly in Physical Autonomy domain after the program. Women (N=24) with schizophrenia had higher QOL than men (N=15) and they improved more after the program compared to men. While SQUALA discriminates between healthy controls and patients, it is not sensitive enough for measuring the effect of psychoeducation. More relevant information about the effect of psychoeducation was obtained from POQ: The patients reported on the importance of delivered information, an opportunity to share the experience with the illness with others, they reconciled better with the fact of being ill and relatives participation in the program was welcomed even required by the patients. The research was supported by CNSLN00B122MSMT CR and Fogarty Program NIH Research Grant # D43 TW05810

Abstract 168

THE IMPACT OF TREATMENT ON THE QUALITY OF LIFE IN DEPRESSED PATIENTS: A COMPARISON OF MIRTAZAPINE AND PAROXETINE

Christine M. Thomas, MSc, Health Economics, Organon Laboratories Ltd, Cambridge, UK, Linda Abetz, MA, Rob Arbuckle, MA, MAPI Values Limited, Macclesfield, Cheshire, UK, Alan G. Wade, PhD, CPS, Community Pharmacology Services, Clydebank, Glasgow, UK

The negative impact of depression on patients' Quality of Life (QoL) is well documented, but treatments of depression have been shown to differ in their impact on QoL. This study assesses differences in the impact of two treatments for depression, Mirtazapine and Paroxetine, on patient reported QoL. A phase IV, prospective, multicentre, randomised, double blind, comparative, 24-week study in primary care was conducted with subjects who had major depressive disorder (DSM IV) (single or recurrent episode(s), resulting in a baseline Hamilton Depression (HAMD) score of >18. QoL was assessed at baseline and weeks 2,4,12 and 24 using the Short Form 36 (SF-36) and the Quality of Life in Depression Scale (QLDS). A total of 197 patients were randomised and received at least one dose of medication (99 for Mirtazapine, and 98 for Paroxetine); 89 patients completed the study (46 for Mirtazapine, and 43 for Paroxetine) and were included in this analysis. Patients were aged 17-74 years (mean 40 years) and 73% of patients were female. Mean QLDS scores improved compared with baseline at all timepoints, in both treatment groups. At all timepoints improvements were greater in the Mirtazapine group (mean change scores range from -0.80 at Week 4 to -14.73 at Week 24) compared with the Paroxetine group (mean change scores range from -0.38 at Week 4 to -10.59 at Week 24). Mean QLDS score at endpoint was significantly lower for the Mirtazapine group as compared to the Paroxetine group (P=0.021). For all SF-36 scales, improvements over time were greater in the Mirtazapine group than the Paroxetine group. At endpoint mean scores for the Mirtazapine group were significantly higher than the Paroxetine group for the SF-36 scales of Physical Functioning (P=0.0003), Pain (P=0.040), Vitality (P=0.046), and Social Functioning (P=0.042). Defining the Minimally Clinically Important Difference (MCID) as 5 points on the SF-36, differences between groups in the mean change scores were therefore clinically meaningful for the scales of Physical Functioning, Pain, Vitality, Social Functioning, and Role-Emotional.

Abstract 169

CAN BUSINESS MANAGERS BENEFIT FROM QUALITY OF LIFE ASSESSMENT AND FEEDBACK

Ciaran A. O'Boyle, PhD, Sarah Clarke, MSc, Psychology, Royal College of Surgeons in Ireland, Dublin 2, UK, Georgina Corscadden, MA, Irish Management Institute, Dublin 16, UK, Lena Ring, PhD, Psychology, Royal College of Surgeons in Ireland, Dublin 2, UK, Stefan Hoefler, PhD, Psychology, Royal College of Surgeons in Ireland, Dublin, UK

Quality of life is an important outcome measure in healthcare but also has significant potential in organisational psychology. This study assessed individual quality of life in company managers and also sought to determine the effects of feedback of QoL information to participants. The SEIQoL was administered, in group settings, by a trained administrator to 107 (51 female) newly appointed managers (NAM) attending management development programmes at a national training agency. The mean age was 34.9 (s.d. = 7.4) years and the mean length of time as a manager was 1.9 (s.d. = 3.1) years. High internal SEIQoL validity (R2 = 0.85, range 0.76-0.92) and reliability scores (r=0.87, range 0.69-0.98) indicated participants understood the task and completed consistent evaluations. The five most important life areas were family (17.6%), work/ career (16.3%), friends/ social life (11%), health (10.3%) and personal relationships (7.3%). The mean SEIQoL Index score was 66.8 (s.d. = 12.0). Compared with previous samples, this was significantly lower than the mean scores for healthy young and elderly adults and for patients with peptic ulcer disease and significantly higher than the mean score for patients receiving palliative care. Participants were provided with feedback on their SEIQoL profiles 2-3 weeks after completion. 83% found the profile to be an accurate reflection of their overall judgement of their QoL and 66% were surprised at some of the outcomes. About half (51%) believed that the SEIQoL findings would alter their view of their lives and 65% said that they would change some aspect of their lives based on the SEIQoL findings. The study demonstrated the applicability of the SEIQoL in measuring the QoL of managers and that the measure can be successfully administered in group settings. Newly appointed managers had significantly lower QoL than previously studied healthy young adults and elderly samples and also patients with peptic ulcer disease. Responses to feedback indicated that the SEIQoL may provide a useful method for personal and organizational development.

Abstract 170

COMPARING AND CONTRASTING WHAT PATIENTS AND HEALTH PROFESSIONALS CONSIDER IMPORTANT TO PERI-OPERATIVE PATIENT QUALITY OF LIFE

Daniel B. Morris, MASc, MEd, Paul C. Hebert, MD, Dean Fergusson, PhD, Jennifer Clinch, MA, Keith G. Wilson, PhD, James Watters, MD, Clinical Epidemiology Program, Ottawa Health Research Institute, Ottawa, ON, Canada

A series of semi-structured interviews was conducted with 33 health professionals and 50 patients undergoing major cardiac, orthopedic, vascular, or (thoracic or abdominal) cancer surgery to compare and contrast how health related quality of life is perceived by patients and health professionals. Health professionals were interviewed once. Patients were interviewed three times, at about one week pre-operatively, within the first week post-operatively and about one month after surgery. Health professional interviews continued until saturation of themes was found and a minimum of two interviews was completed from each professional group, this occurred after 33 interviews. Patient interviews continued until saturation and at least 10 post-operative interviews were completed within each surgical group, this occurred after approximately 115 interviews. Themes related to peri-operative patient quality of life were extracted from health professional and patient interviews using the principles of grounded theory with NVivo 2.0 software. Health professionals and patients indicated many themes in common including anxiety, familial support, pain, positive outlook, practical help, and uncertainty. However, health professional were more likely than patients to indicate themes such as complications, delays, prognosis, survival, and return to work. Whereas, patients were more likely than health professional to indicate a few issues including faith, and planning ahead. Although many similarities were found between patient and health professional interviews, differences were found. These differences are important in reminding researchers and clinicians alike that quality of life is a subjective construct and that existing measures, often created with limited or no patient input, may omit important elements of great significance to patients.

Abstract 171

THE DEVELOPMENT OF THE MODEL FOR PERI-OPERATIVE PATIENT QUALITY OF LIFE

Daniel B. Morris, MASc, MEd, Paul C. Hebert, MD, Dean Fergusson, PhD, Jennifer Clinch, MA, Keith G. Wilson, PhD, James Watters, MD, Clinical Epidemiology Program, Ottawa Health Research Institute, Ottawa, ON, Canada

A series of semi-structured interviews was conducted with 33 health professionals and 50 patients undergoing major cardiac, orthopedic, vascular, or (thoracic or abdominal) cancer surgery to develop a comprehensive model of peri-operative patient quality of life that will form the basis of a new assessment tool. Health professionals were interviewed once. Patients were interviewed three times, at about one week pre-operatively, within the first week post-operatively and about one month after surgery. Health professional interviews continued until saturation of themes was found and a minimum of two interviews was completed from each professional group, this occurred after 33 interviews. Patient interviews continued until saturation and at least 10 post-operative interviews were completed within each surgical group, this occurred after approximately 115 interviews. Themes relating to peri-operative patient quality of life were extracted from the interviews using the principles of grounded theory with NVivo 2.0 software. These themes were refined and structured to develop a meaningful model for peri-operative patient quality of life. Five main domains were extracted: physical, social, emotional, spiritual, and cognitive well-being. Within each of these major domains a hierarchy of more detailed themes was established. The model developed here based on over 100 interviews with surgical patients and health professionals can be used to provide a rationale for quality of life definition and instrument selection for the peri-operative patient population, thereby increasing the relevance and validity of the results.

Abstract 172

ASSESSING MEASUREMENT EQUIVALENCE ACROSS UPPER LIMB SUBGROUPS

Anjali Mazumder, BSc(Hons), Data Tools and Statistical Methods, Dorcas E. Beaton, PhD, Measurement of Health & Function, Institute for Work & Health, Toronto, ON, Canada

Attaining measurement equivalence across subpopulations is an important psychometric and clinical trait in assuring the quality of a health outcome measure. The Disabilities of the Hand, Arm and Shoulder (DASH) Outcome Measure

is a self-completed whole extremity measure that was developed to determine the impact of a disorder from the perspective of the whole upper limb. Several studies have shown the instrument to be a quality outcome measure based on the traditional psychometric concepts of reliability, validity and responsiveness. Although the DASH demonstrates within population validity, there is little evidence to show that the instrument is comparable across different subgroups within a population. The purpose of this study was to assess measurement equivalence across upper limb subgroups of the DASH outcome measure. The multi-center study sample consisted of 200 patients who can be dichotomized into having either shoulder or wrist/hand musculoskeletal disorders. Both the single parameter Rasch model and the two parameter partial credit model of Samejima's was employed to calibrate the responses. Muraki's stepwise differential item functioning (DIF) procedure was used to detect systematic shifts in measurement performance across the shoulder and wrist/hand subpopulations. Results of the DIF analysis indicate that there are differences in the behavior of the two subgroups. For instance, preliminary results suggest that the shoulder patients are more likely to indicate an ability to open a tight jar unlike its wrist counterpart. Conversely, shoulder patients are less likely (15.7%) to indicate a disability level wrist/hand patients. Approximately one-third of the items exhibit differential item functioning for upper-limb subgroups. Although this substantiates the concern that the DASH does not measure the performance of the upper-limb subgroups equivalently, it does not invalidate the quality of the outcome measure. It does suggest, however, that the cross-comparison of upper limb subgroups requires determining the systematic shift in response categories thresholds in order to understand the behavioral differences.

Abstract 173

PROGNOSIS OF UPPER-EXTREMITY SOFT TISSUE DISORDERS

Anusha Govinda Raj, MSc, John Frank, MD, Epidemiology of Disability, Sheilah Hogg-Johnson, PhD, Data Tools and Statistical Methods, Institute for Work & Health, Toronto, ON, Canada, Elizabeth Badley, PhD, Public Health Sciences, University of Toronto, Toronto, ON, CANADA

In workers with musculoskeletal injuries of the neck and upper limb it is estimated that 7% of claims account for 75% related costs, usually in workers with prolonged course of recovery. The benefit of being able to identify these workers early in the course is evident. Usually, models involve duration of time off work, which may be an administrative rather than health outcome. This study models the prognosis of workers with upper-extremity disorders using pain, function and role limitations as outcome measures. The objective is to identify factors associated with a prolonged course. 403 workers with accepted lost time WCB claims participated in a longitudinal (52 week) telephone survey. Data used in this analysis was that gathered at baseline, 4 and 16 weeks. Dependent variables were Von Korff's pain grade (modified for acute injuries) (PG), the modified American Shoulder and Elbow Surgeons scale (ASES) and the Short-Form-36 role-physical subscale (RFP). Independent variables included demographic, symptoms/function, nature of workplace, workplace-psychosocial and expectations. Linear (PG, ASES) or logistic (RFP) regression identified significant univariate predictors ($p < 0.25$) which then entered the final multivariable model and was reduced using backward manual elimination ($p < 0.05$). The 3 prediction models explained 24% (PG), and 43% (ASES, RFP) of the variance respectively. Higher pain levels at 16 weeks were associated with a smaller difference in pain levels between baseline and 4 weeks and female gender. Worse function at 16 weeks was associated with lower earnings and higher perceived risk of re-injury. Role limitations at 16 weeks was associated with presence of other health conditions and higher perceived risk of re-injury. Recovery expectations and function at 4 weeks were common predictors of all 3 models. Pessimistic recovery expectations and worse function at 4 weeks were associated with higher pain levels, worse function and presence of role limitations at 16 weeks. Patient perceptions are important in determining health outcomes and insightful in understanding recovery. outcome. This study models the prognosis of workers with upper-extremity disorders using pain, function and role limitations as outcome measures. The objective is to identify factors associated with a prolonged course. 403 workers with accepted lost time WCB claims participated in a longitudinal (52 week) telephone survey. Data used in this analysis was that gathered at baseline, 4 and 16 weeks. Dependent variables were Von Korff's pain grade (modified for acute injuries) (PG), the modified American Shoulder and Elbow Surgeons scale (ASES) and the Short-Form-36 role-physical subscale (RFP). Independent variables included demographic, symptoms/function, nature of workplace, workplace-psychosocial and expectations. Linear (PG, ASES) or logistic (RFP) regression identified significant univariate predictors ($p < 0.25$) which then entered the final multivariable model and was reduced using backward manual elimination ($p < 0.05$). The 3 prediction models explained 24% (PG), and 43% (ASES, RFP) of the variance respectively. Higher pain levels at 16 weeks were associated with a smaller difference in pain levels between baseline and 4 weeks and female gender. Worse function at 16 weeks was associated with lower earnings and

higher perceived risk of re-injury. Role limitations at 16 weeks was associated with presence of other health conditions and higher perceived risk of re-injury. Recovery expectations and function at 4 weeks were common predictors of all 3 models. Pessimistic recovery expectations and worse function at 4 weeks were associated with higher pain levels, worse function and presence of role limitations at 16 weeks. Patient perceptions are important in determining health outcomes and insightful in understanding recovery.

Abstract 174

INTERPRETING THE CLINICAL RELEVANCE OF DIFFERENCES IN PATIENT-REPORTED OUTCOMES (PROS) IN PATIENTS TREATED FOR REFLUX DISEASE

Ola Junghard, PhD, Biostatistics, Ingela Wiklund, PhD, Outcomes Research, AstraZeneca, Moelndal, Sweden

To explore what can be considered to be a clinically relevant treatment difference in PROs with a 7-graded response format by relating it to differences in rates of symptom resolution. PRO data from a clinical trial (n=997), comparing omeprazole 20 mg with omeprazole 10 mg and ranitidine 150 mg bid in patients with reflux disease, were used in this analysis. The PRO data were collected using the Gastrointestinal Symptom Rating Scale (GSRS) Reflux dimension, which consists of two 7-graded items, one for heartburn and one for acid regurgitation. The mean change from baseline to 4 weeks is compared with the proportion of patients being symptom-free (i.e., score=None on both items in the GSRS reflux dimension) after 4 weeks of treatment. A treatment difference of 10 %-points is often considered as clinically relevant and is used as an anchor. The mean Reflux dimension score change was 1.48 for omeprazole 20 mg, 1.32 for omeprazole 10 mg and 1.15 for ranitidine. The corresponding proportion of patients with complete symptom relief during the past 7 days was 42%, 34% and 26%, respectively. Thus the difference between omeprazole 20 mg and ranitidine 150 mg of 0.33 score points corresponds to a difference in the proportions of symptom-free patients of 16%. The difference between omeprazole 20 mg and omeprazole 10 mg of 0.16 score points corresponds to a difference in proportions of 8%. Hence, a treatment difference of 10 %-points corresponds to a difference in mean Reflux dimension score of 0.2. The same relationship – that a difference of 10 % points corresponds to a difference in mean score of 0.2 – was found when evaluating the GSRS heartburn item separately. The results suggest that a treatment difference of 0.2 score points defines a clinically relevant treatment difference, at least in studies where a large proportion of the patients become symptom-free.

Abstract 175

THE IMPACT OF BASELINE SCORES ON A RELEVANT CHANGE AND A MINIMAL IMPORTANT DIFFERENCE (MID) IN PATIENT-REPORTED OUTCOMES (PROS) IN REFLUX DISEASE

Ingela Wiklund, PhD, Outcomes Research, Ola Junghard, PhD, Biostatistics, AstraZeneca, Moelndal, Sweden

Data from 1774 endoscopy -ve patients with mild to severe heartburn were analyzed to study the impact of baseline scores on changes and the minimal difference between treatments in PROs with a 7-point scale, the GSRS Reflux dimension, and the Sleep and Eat/drink dimensions in QOLRAD. The OTE was used to mirror the clinical relevance of the change and to establish the MID. The magnitude of the PRO change after 4 weeks of esomeprazole treatment differed considerably depending on the baseline score. The mean change in the Reflux dimension in patients with severe symptoms (n=361) was 3.5, the mean change in the Sleep dimension in patients with severe problems (n=308) was 3.3, and the mean change in the Eat/drink dimension in patients with severe problems (n=399) was 3.3. The corresponding changes in patients with none or minor symptoms/problems were 0.7, 0.4, 0.6 for Reflux (n=316), Sleep disturbances (n=638) and Eat/drink problems (n=298), respectively. In patients with moderate symptoms or problems the change scores were 2.5, 2.4 and 2.3 for Reflux (n=541), Sleep (n=389) and Eat/drink (n=570) respectively. Irrespective of baseline severity, around 80% of the patients considered their symptoms to be “moderately better” according to OTE. The MID patients with severe symptoms/problems are estimated to be 2-5 times the MID for patients with no or minor symptoms/problems. For both QOLRAD dimensions, a MID was estimated to 0.2 for patients with none/minor baseline problems, to 0.4, 0.6 and 0.9 for those with mild, moderate and severe baseline problems, respectively. For the Reflux dimension the MID increased from 0.3 to 0.9 when the severity of baseline symptoms increased from minor to severe. Hence, a 0.5 score change on a 7-point scale may represent a change perceivable by patients, but a difference of 0.5 score points represents an unrealistic target when comparing two treatments. Hence, the baseline score must be considered when interpreting data.

Abstract 176

VALIDATION OF THE REFLUX DISEASE QUESTIONNAIRE (RDQ), A SYMPTOM SCALE FOR USE IN PATIENTS WITH UPPER GASTROINTESTINAL (GI) SYMPTOMS

Ingela Wiklund, PhD, Outcomes Research, AstraZeneca, Molndal, Sweden, Sander Veldhuyzen van Zanten, PhD, Health Science, Queen Elisabeth II Health Sci Center, Halifax, NS, Canada, David Armstrong, MA, Gastroenterology, Hamilton Health Science Center, Hamilton, ON, Canada, Alan Barkun, MD, McGill University Health Centre, Montreal, ON, Canada, Robert White, PhD, Bijan Chakraborty, MStat, Lisa Tanser, MSc, Krista Nevin., AstraZeneca Canada, Mississauga, ON, Canada

Patients with heartburn-dominant (HB 35%) or non-heartburn dominant (NHB 65%) uninvestigated upper GI symptoms from the Confirmatory Acid Suppression Test study were used to validate RDQ. The RDQ has four dimensions (6-point scale); heartburn, regurgitation, epigastric pain, and GERD. 1121 patients (49,5% male; mean age 45.9 years) were randomised to esomeprazole 40 mg, either o.d. or b.i.d. for 7 days, followed by o.d. treatment for 3 weeks. The RDQ, the Global Overall Symptoms (GOS), and the QoLRAD were completed at baseline and after 4 weeks of treatment. Construct validity of the RDQ was measured by the Spearman correlation coefficient (SCC) of changes from baseline to 4 weeks between RDQ with the QoLRAD, and the GOS. Responsiveness was assessed by calculation of the effect size ($ES = DRDQ / SD$ at baseline). The test-retest reliability of the RDQ was assessed using the intraclass correlation coefficient (ICC) for patients with stable disease over 4 weeks. The Cronbach alpha assessed internal consistency reliability. Table 1 shows internal consistency, test retest reliability and effect size by NHB and HB groups. The RDQ is a valid, reliable and responsive instrument in both HB and NHB dominant patients with uninvestigated upper GI symptoms and is useful in clinical trials.

Abstract 177

THE USE OF BOOTSTRAP METHODS FOR ESTIMATING SAMPLE SIZE FOR HEALTH-RELATED QUALITY OF LIFE OUTCOMES

Stephen J. Walters, PhD, Michael J. Campbell, PhD, ScHARR, University of Sheffield, Sheffield, UK

HRQoL outcomes tend to generate data that have discrete, bounded and skewed distributions. Thus, standard methods of sample size estimation may not be appropriate. For this reason, non-parametric methods can be used to estimate sample sizes for HRQoL outcomes. The bootstrap is one such non-parametric method for estimating sample sizes. We describe and compare four different methods for estimating sample size and power, when the main outcome is a HRQoL measure. These methods are: 1. assuming a Normal distribution and comparing two means; 2. using a non-parametric method; 3. Whitehead's method based on the proportional odds model; 4. the bootstrap. We illustrate the various methods, using data from the SF-36. The results show that if the HRQoL outcome has a limited number of discrete values (< 7) and/or the expected proportion of cases at the boundaries is high (scoring 0 or 100), then we would recommend using Whitehead's method. Alternatively, if the HRQoL outcome has a large number of distinct values and the proportion at the boundaries is low, then we would recommend using method 1. If a pilot or historical dataset is readily available (to estimate the shape of the distribution) then bootstrap simulation (Method 4) based on this data will provide a more accurate and reliable sample size estimate than conventional methods. In the absence of a reliable pilot set, bootstrapping is not appropriate and conventional methods of sample size estimation will need to be used. Fortunately, with the increasing use of HRQoL outcomes in research, historical datasets are becoming more readily available. Strictly speaking, our results and conclusions only apply to the SF-36

outcome measure. Further empirical work is required to see whether these results hold true for other HRQoL outcomes. However, the SF-36 has many features in common with other HRQoL outcomes, such as the QLQ-C30, e.g. multi-dimensional, ordinal or discrete response categories with upper and lower bounds, and skewed distributions. Therefore, we believe these results and conclusions using the SF-36 may be appropriate for other HRQoL measures.

Abstract 178

TRANSITION RATING BIAS: OVERESTIMATE OF IMPROVEMENTS IN CLINICAL STATUS BY CLINICIANS
Ann-Christin Mörk, PhD, Elisabeth Ståhl, PhD, Clinical Science, AstraZeneca R&D, LUND, Sweden, Elizabeth F. Juniper, Prof, McMaster University, HAMILTON, ON, Canada

The original purpose of this study was to determine the Minimal Important Difference for the Asthma Control Questionnaire (ACQ), a validated instrument with strong measurement properties (1). However, an unexpected tendency of clinicians to overestimate improvements in asthma control thwarted the endeavour. We describe the observed clinician bias and discuss its implications for clinical studies. 94 adults with inadequately controlled asthma received a full clinical consultation with one of nine asthma specialists. Medications were adjusted according to clinical needs. Four weeks later the same clinician estimated change in asthma control on a 15-point scale (-7 = a very great deal worse, 0 = no change, +7 = a very great deal better). Patients completed the ACQ before each consultation but responses were not shown to the clinician. Clinicians consistently recorded that patients improved more than their change in ACQ scores suggested ($p < 0.05$). Possible explanations for this bias include: 1) patients want to please their doctor, 2) clinicians expect patients treated with efficacious interventions to improve, 3) transition ratings may reflect current asthma control more closely than change (2), 4) knowledge of the patients' status prior to baseline may have influenced clinicians' estimate of change. In conclusion, investigators should be aware of potential biases that may occur when clinicians are asked to estimate change in clinical status using transition rating scales compared with measuring absolute status at each visit. (1) Juniper et al. Development and validation of a questionnaire to measure asthma control. *Eur Respir J* 1999; 14: 902-7. (2) Guyatt et al. A critical look at transition ratings. *J Clin Epidemiol* 2002; 55: 900-908.

Abstract 179

MEASURING DECLINES IN HEALTH-RELATED QUALITY OF LIFE IN OLDER PERSONS
Jennifer D. Walker, BSc, Jenny X. Zhang, MPA, Colleen J. Maxwell, PhD, Community Health Sciences, University of Calgary, Calgary, AB, Canada, Walter P. Wodchis, PhD, Adriana Venturini, MSc, Toronto Rehabilitation Institute, Queen Elizabeth Centre, Toronto, ON, Canada, David B. Hogan, MD, Community Health Sciences, University of Calgary, Calgary, AB, Canada, David H. Feeny, PhD, Pharmacy and Pharmaceutical Sciences, University of Alberta, Edmonton, AB, Canada

The measurement of health-related quality of life (HRQL) provides an opportunity to assess changes in health status and care outcomes among older persons with diverse medical conditions. This study aims to validate the Health Utilities Index-Mark2 (HUI2) and the Minimum Data Set-Health Status Index (MDS-HSI) for older persons by investigating the extent to which declines in HRQL relate to declines on other measures of health status. A sample of 72 older (65+) high-risk home care clients participating in an evaluation of a community care program in Calgary, Canada was included in the analyses. All subjects received MDS-Home Care and HUI2 assessments at baseline and 12-month follow-up to calculate the HRQL measures. Multiple logistic regression was used to examine correlates of HRQL decline (drop of 0.03 points or more on the HUI2 or MDS-HSI over 12 months), including age, sex, declines in activities of daily living (ADL) and cognition, and increased comorbidity, frailty, polypharmacy, and depression. Over 12 months, the sample experienced increased polypharmacy and frailty and decreased ADL function. These variables, along with age and sex, were retained for further analyses. Following adjustment for other covariates, clinically significant MDS-HSI decline was not significantly associated with declines on other health and functional status measures, although all relationships were in the expected direction. Decline in ADL function approached significance ($p = 0.054$). Declines on the HUI2 were accompanied by a significant increase in frailty (OR=3.2, 95%CI:1.1-9.4). Though not statistically significant, the results showed a negative association between HUI2 decline and increased age and polypharmacy. While previous research has shown that the HUI2 and MDS-HSI accurately reflect the health of older populations, these longitudinal results show that declines in these scores

did not accompany change in other measures of health status. Although the findings may reflect limitations due to small sample size or follow-up time, they suggest a lack of concurrent validity for these HRQL measures in older populations, particularly the HUI2.

Abstract 180

A COMPARISON OF CHANGES IN OBSERVED AND SELF-REPORTED HRQL IN COMMUNITY-BASED SENIORS

Jenny X. Zhang, MPA, Jennifer D. Walker, BSc, Colleen J. Maxwell, PhD, Community Health Sciences, University of Calgary, Calgary, AB, Canada, Walter P. Wodchis, PhD, Adriana Venturini, MSc, Toronto Rehabilitation Institute, Queen Elizabeth Centre, Toronto, ON, Canada, David B. Hogan, MD, Community Health Sciences, University of Calgary, Calgary, AB, Canada, David H. Feeny, PhD, Pharmacy and Pharmaceutical Sciences, University of Alberta, Edmonton, AB, Canada

Two measures of HRQL were examined in this study: (i) a self-reported measure ~ the Health Utilities Index Mark 2 (HUI2) and (ii) an observed measure ~ the Minimum Data Set-Health Status Index (MDS-HSI). The objectives were to compare changes and clinically significant declines in the 2 HRQL measures over one year and to examine the relative contribution of changes in single-domains to changes in overall HRQL. A sample of 72 high-risk home care clients (65+) attending a comprehensive community care program in Calgary, Canada were included in the analyses. All subjects were assessed with the MDS for Home Care v2.0 and HUI2 assessments at baseline and 12-month follow-up times. A paired t-test was used to compare the changes in the 2 measures. Multivariate linear regression models were used to examine the relative importance of changes in single-domains to changes in overall HRQL. Clinically significant declines were defined as a decrease of at least 0.03 and 0.05 for total scores and single-domain scores, respectively. Agreement between the 2 HRQL measures in clinically significant declines for total and domain scores were calculated with McNemar's test. The mean change in HUI2 (-0.022, sd=0.21) did not differ significantly from that observed with the MDS-HSI (-0.027, sd=0.21). Mean changes in domain scores were also comparable for the 2 measures. Regression analyses showed that changes in the pain domain explained most of the changes in total scores for both HUI2 and MDS-HSI. Changes in the sensation and self-care domains were significantly associated with changes in the HUI2 but not with the MDS-HSI. The proportion of subjects with clinically significant declines, for either total or domain scores, were comparable for both HRQL measures. The results indicate that the observed and self-reported HRQL measures showed comparable changes over one year. The findings for the pain domain suggest the importance of pain management to overall quality of life among older community-based high-risk seniors.

Abstract 181

HEALTH STATUS OF THE GENERAL ADULT US POPULATION AS ASSESSED BY THE EQ-5D AND HEALTH UTILITIES INDEX

Nan Luo, PhD, Jeffrey A. Johnson, PhD, Public Health Sciences, University of Alberta, Edmonton, Ab, Canada, James W. Shaw, MS, Pharmaceutical Sciences, University of Arizona, Tucson, AZ, David H. Feeny, PhD, Institute of Health Economics, University of Alberta, Edmonton, AB, Canada, Stephen J. Coons, PhD, Pharmaceutical Sciences, University of Arizona, Tucson, AZ

Objective: The purpose of this analysis was to compare the self-reported health status of the adult US population using three different multi-attribute preference-based measures, the EQ-5D and Health Utilities Index Mark 2 (HUI2) and Mark 3 (HUI3). Methods: We conducted a national survey of the community-dwelling adult US population, with oversampling of the two largest minority groups, Hispanics and non-Hispanic blacks. Respondents to the in-home interviews self-completed the EQ-5D and HUI2/3 questionnaires. Rates of health problems and mean (se) overall health index scores of the adult US population and its subgroups were estimated and construct validity of these measures was assessed by testing a priori hypotheses that related these measures to each other. Results: A total of 4,048 interviews were completed (overall response rate: 59.4%). The sample was predominantly female (58.2%), with a mean (se) age of 43.0 (0.3) years; 12.6% of the sample was 65 years of age or older. It was estimated that 15.2% (HUI3) to 47.9% (EQ-5D) of the adult US population was in full health (i.e. no reported health problems). Estimated mean (se) overall index scores (where 1.0 = perfect health and 0.0 = dead) for the adult US

population as assessed by the EQ-5D, HUI2 and HUI3 were 0.84 (0.01), 0.86 (0.01) and 0.81 (0.01), respectively. The three overall preference scores were strongly inter-correlated ($\rho > 0.60$). Correlations between attributes from different measures ranged from 0.10 to 0.67. As expected, similar attributes (e.g., EQ-Mobility and HUI3 Ambulation, EQ-Pain/Discomfort and HUI3 Pain) were strongly correlated ($\rho > 0.60$), while dissimilar attributes (e.g., EQ-Anxiety/Depression and HUI2 Sensation) were weakly correlated ($\rho < 0.20$). Discussion: This study provides nationally representative estimates of self-reported health status for the EQ-5D, HUI2 and HUI3 for the adult US population. Similarities in overall index scores may mask differences between the measures in attributes/dimensions and the distributions of the overall scores.

Abstract 182

CHRONIC VENOUS DISEASE: THROUGH BODY MASS INDEX

Jean Jérôme Guex, Pr, Société Française de Phlébologie, Nice, France, Eric Myon, PhD, Charles Taieb, MD, Public Health, Health Economics & Quality of life, I.R.P.F., Boulogne Billancourt, France, Robert Launois, Pr, REES -, Réseau d'Evaluation en Economie de la Santé, Paris, France

Introduction: Many studies have confirmed obesity as a Chronic Venous Disease (CVD) risk factor. Few studies have described the pathology through Body Mass Index (BMI). Objective: To describe the impact of overweight in CVD. Method: Between May and July 2003, 567 GP have recruited 1049 female patients spontaneously consulting for CVD. The patients could fill in a series of validated questionnaires in order to evaluate the consequences of their disease. Results: The results of the study concern 1045 patients with a mean age of 44,45 years old (± 10.70) (min: 18 -max: 65) ; 66% have a professional activity. The patients average size was 164.39 cm (± 5.99) for an average weight of 65.2 kg (± 12.5). The BMI calculation gives an average BMI of 24.17 (± 4.71). The values issued by the WHO have been taken into account: Thinness: 4% -Normal weight 62% - Overweight: 24% - Obesity: 10%. For each of these subgroups, CIVIQ score is respectively of 21.2, 16.6, 25.8, 32.1. In order to make easier the analysis, we have reduced to two subgroups BMI < 27 vs > 27. CIVIQ score is: 29,8 versus 40,9 ($p < 0.0001$). This difference is found through the severity (CEAP) classification: 15% of C0-C2 have a BMI > 27, while they represent 26% of the C3-C6 ($p < 0.001$). We have tested both subgroups on sedentary lifestyle, family history, underfloor heating, pregnancy risk factors. None of them is significant except sedentary lifestyle (61% vs 76%, $p < 0.0001$) Conclusion : A more important CVD severity grade is expected for a BMI > 27.

Abstract 183

CHRONIC VENOUS DISEASE: PATIENTS PROFILE

Lionel Didier, PhD, Public Health, Pierre Fabre Santé, Castres, France, Charles Taieb, MD, Eric Myon, PhD, Nathalie Marionneau, PhD, Public Health, Health Economics & Quality of life, I.R.P.F., Boulogne Billancourt, France, Claire Nguyen, MD, Medical Direction, Pierre Fabre Santé, Castres, France, Jean Jérôme Guex, Pr, Société Française de Phlébologie, Nice, France

Background: Chronic venous disease (CVD), (including pain, sensation of swelling, cramps, varices) is a frequent condition which also results in important socio-economic repercussions. Its prevalence in the French adult population has been estimated at 57% for women in 1996. Although chronic venous disorders represent an important public health issue, until recently the patients' profile of venous disease sufferers was poorly understood. Objective :The objective of our study was to describe the profile of French women suffering from CVD. Method: Symptomatic women patients suffering from CVD (CEAP clinical classes C0 to C4), aged over 18, whether or not wearing compression stockings, newly treated by their GP with a phlebotropic drug were enrolled in the study. Every patient had to complete a self-questionnaire including the SF-12, the CES-D, the Epworth and the CIVIQ scales at day 0, day 3 and day 7. Results: This analysis includes the first 399 patients assessed at inclusion. The mean age was 45.0 years old (SD=11, n=370). 65.7% have a professional activity, 32% practice sport and 33% are smokers. 78.4% gave birth already and 50% are under oral contraceptive. 9.1% wear compression stockings. (see Table 1) Conclusion: These results demonstrate that CVD has a great impact on women. The SF-12 mean scores were below those of the age- and gender-matched general population. Women with CVD report greater risk of high depressive symptoms, compared to the study of Rield where 23.1% of women reported high depressive symptoms (CES-D score ≥ 16 , age

between 65 and 75). The impact of CVD on patients daily life is high even if it seems relative compared to the mean scores obtained by Launois when initially validating the CIVIQ; for example for patients suffering venous insufficiency of lower limb and arteritis mean score was 53.08(SD=14.9), unfortunately comparison data with patients suffering CVD are lacking. Concerning the impact of CVD on excessive daytime sleepiness, it is huge compared to the prevalence of sleep disorders in an employed Swiss population (n= 668; EDS: 13% - Schmitt BE et al).

Abstract 184

CHRONIC VENOUS DISEASE: DEPRESSIVE SYMPTOMATOLOGY

Eric Myon, PhD, Charles Taieb, MD, Public Health , Health Economics & Quality of life, I.R.P.F., Boulogne Billancourt, France, Jean Jérôme Guex, Pr, Société Française de Phlébologie, Nice, France

Background: Chronic venous disease (CVD), (including pain, sensation of swelling, cramps, varices) is a frequent condition which also results in important socio-economic repercussions. Its prevalence in the French adult population has been estimated at 57% for women in 1996. Although chronic venous disorders represent an important public health issue, studies on its impact on daily life and specifically psychological suffering are scarce. Objective: The objective of our study was to assess depressive symptomatology (DS) among CVD affected women. Methods :Symptomatic women patients suffering from CVD (CEAP clinical classes C0 to C4), aged over 18, whether or not wearing compression stockings, newly treated by their GP with a phlebotropic drug were enrolled in the study. Every patient had to complete a self-questionnaire including the CES-D scale at day 0, day 3 and day 7. A score at or above 17 indicates a possible DS, a score at or above 23 indicates a probable DS. Results :This analysis includes the first 371 patients assessed at day 0, D3 and D7. The mean age was 45.0 years old (SD=11, n=370). The mean CES-D scores at day 0, D3 and D7, were respectively 14.9 (SD=10.2), 13.7 (SD=8.9) and 12.8 (SD=10.1). The results highlight a possible DS in our population (score \geq 17) for 36.3%, 32.3% and 29.0% respectively at day 0, D3 and D7 (p< 0,01, n=328,). Focusing on patients that have expressed a probable DS, they were 74 at inclusion (22.0% of the population); they show a significant improvement of their status assessed by CES-D. From those 74 patients, only 50 still had a score \geq 23 at D3 and 46 at D7 showing a decrease of 37.8% of the number of patients expressing a probable DS (p< 0,0001, n=74, matched test J0-J7). Conclusion: It is recognised that women report greater risk of high depressive symptoms. In the study of Rield assessing depressive symptoms in older women (age between 65 and 75), 23.1% of women reported high depressive symptoms (CES-D score \geq 16). Those results compared to ours demonstrate that CVD result in psychological effects that seriously affect patients lives.

Abstract 185

CHRONIC VENOUS DISEASE: HEALTH STATUS

Jean Jérôme Guex, Pr, Société Française de Phlébologie, Nice, France, Claire Nguyen, MD, Medical Direction, Pierre Fabre Santé, Castres, France, Eric Myon, PhD, Nathalie Marionneau, PhD, Charles Taieb, MD, Public Health , Health Economics & Quality of life, I.R.P.F., Boulogne Billancourt, France

Background: Chronic venous disease (CVD), (including pain, sensation of swelling, cramps, varices) is a frequent condition which also results in important socio-economic repercussions. Its prevalence in the French adult population has been estimated at 57% for women in 1996. Although chronic venous disorders represent an important public health issue, studies on its impact on quality of life are scarce. Objective : The objective of our study was to assess health status among women suffering from CVD. Methods : Symptomatic women patients suffering from CVD (CEAP clinical classes C0 to C4), aged over 18, whether or not wearing compression stockings, newly treated by their GP with a phlebotropic drug were enrolled in the study. Every patient had to complete a self-questionnaire including the SF-12 scale at day 0, day 3 and day 7. The SF-12 is a generic measure

of health status. It can be self-administered to most people in 2 minutes or less and has been used with high degree of acceptability and data quality. The SF12 is composed of 2 dimensions, a Physical Component Summary (PCS-12) and a Mental component Summary (MCS-12). The results are standardised on the general US population (mean score of 50 (SD=10)) so results for one can be meaningfully compared with the other. The lower the score is the worst is the impact on patients_ quality of life. Results: This analysis includes the first 399 patients assessed at day 0, day 3 and day 7. The mean age was 45.0 years old (SD=11, n=370). At inclusion time (n=374), MCS-12 and PCS-12 were respectively 44.7 (SD=10.6) and 46.4 (SD=8.4); at day 3 and day 7, these dimensions were respectively: D3: 46,5 (SD=10,2) and 46.2 (SD=8.0) D7: 48.0 (SD=10.3) and 46.2 (SD=7.8). For the mental dimension, the difference was statistically significant (p=0.0001). Conclusion: These results suggest that CVD has a great impact on women. The SF-12 mean scores were below those of the age- and gender-matched general population. The patient management and the use of a phlebotropic drug demonstrated an improvement on the mental health status of the patient and a decrease of the impact of pain interfering with patients_ normal work.

Abstract 186

CHRONIC VENOUS DISEASE: & SPECIFIC QUALITY OF LIFE

Robert Launois, Pr, REES, Réseau d'Evaluation en Economie de la Santé, Paris, France, Eric Myon, PhD, Charles Taieb, MD, Nathalie Marionneau, PhD, Public Health , Health Economics & Quality of life, I.R.P.F., Boulogne Billancourt, France, Lionnel Dider, PhD, Health, Pierre Fabre Santé, Castres, France, Jean Jérôme Guex, Pr, Société Française de Phlébologie, Nice, France

Background: Chronic Venous Disease (CVD) may have a substantial effect on functioning and may seriously affects patients' quality of life (QoL). Its prevalence in the French adult population has been estimated at 57% for women in 1996. Although chronic venous disorders represents an important public health issue, studies on its impact on QoL are scarce. Objective: The objective of our study was to assess specific quality of life among women suffering from CVD. Method: Symptomatic women patients suffering from CVD (CEAP clinical classes C0 to C4), aged over 18, whether or not wearing compression stockings, newly treated by their GP with a phlebotropic drug were enrolled in the study. Every patient had to complete a self-questionnaire including the CIVIQ scale at day 0, day 3 and day 7. Results: This analysis includes the first 399 patient who completed the questionnaires at day 0, D3 and D7. For all the dimensions, a score decrease between day 0 and day 3 and then between day 3 and day 7 was observed . Scoring, on average, 44,7 at day 0, the Pain dimension is the one which consequences are the most uncomfortable for patients. Psychical repercussions seem to be less important than physical and social ones. Dimensions and total scores reduction has been observed for mean scores as well: this means patients quality of life improvement regarding the 4 dimensions and the global score as well and is statistically significant between D0 and D3, D0 and D7 and D3 and D7. Friedman test has been used to simultaneously comparing the 3 periods as a whole: this global test confirms the QoL scores decrease reflecting a QoL improvement between day 0, D3 and D 7 (p =0.0001). Conclusion: There was evidence of an improvement in QoL following patient management and the use of a phlebotropic drug. This improvement was very effective despite the relative impact of CVD on our patients' QoL, compared to the mean scores obtained by Launois when initially validating the CIVIQ; for example for patients suffering venous insufficiency of lower limb and arteritis mean score was 53.08(SD=14.9).

Abstract 187

CHRONIC VENOUS DISEASE: CARE IMPACT

Charles Taieb,MD, Public Health , Health Economics & Quality of life, I.R.P.F., Boulogne Billancourt, France, Lionnel Didier, PhD, Health, Pierre Fabre Santé, Castres, France, Eric Myon, PhD, Public Health , Health Economics & Quality of life, I.R.P.F., Boulogne Billancourt, France, Robert Launois, Pr, REES, Réseau d'Evaluation en Economie de la Santé, Paris, France, Jean Jérôme Guex, Pr, Société Française de Phlébologie, Nice, France

Introduction: The CVD treatment is based on a double treatment, either conventional (contention or venotonics) nor radical (sclerotherapy, surgery). **Objective:** Describing the venotonic and contention association impact on the patients quality of life. **Method:** Between May and July 2002, 567 GP have recruited 1045 female patients spontaneously consulting for CVD. Two patients subgroups have been identified: RA (treated with "ruscus aculeatus, hespéridine méthyl chalcone HMC & acide ascorbique Vit.C), RAC: (treated with RA and contention). **Results:** In both subgroups RA (n=697) and RAC (n=269), the risk factors have been compared: sedentary lifestyle, family history, underfloor heater, pregnancy. Obesity and family history were found most often among the RAC patients (25% v 16% and 50% v 34% , p<0.001 ki2). At inclusion, the specific (CIVIQ), non specific (SF12) quality of life (QoL) and daytime sleepiness (Epworth scale) were evaluated through a self-questionnaire. 304 patients have answered at D0 and D7. No significant difference has been observed between the 2 groups RAC v RA; CIVIQ:32.3 v 32.3 , SF12 : Physical dimension: 45 v 46.9, Mental dimension: 43.7 v 45, Epworth: 8.4 v 7.5. After a 7 day treatment, the same scales were administered. In the RAC group, CIVIQ has improved (p=0.0004). In the RA group SF12 Mental Dimension, CIVIQ and Epworth significantly improved at D7 versus D0 with respectively p<0.001, =0.036, p<0.001). **Conclusion :** In the chronic venous disease, associating a contention to a venotonic prescription do not improve the patient's quality of life.

Abstract 188

CHRONIC VENOUS DISEASE: COMPLIANCE WITH TREATMENT

Eric Myon, PhD, Public Health, Health Economics & Quality of life, I.R.P.F., Boulogne Billancourt, France, Robert Launois, Pr, REES, Réseau d'Evaluation en Economie de la Santé, Paris, France, Jean Jérôme Guex, Pr, Société Française de Phlébologie, Nice, France, Charles Taieb, MD, Public Health , Health Economics & Quality of life, I.R.P.F., Boulogne Billancourt, France

Introduction: The compliance with the treatment has often been studied for pathologies such as the high blood pressure or hypercholesterolemia; paradoxically, it has been very little studied for chronic venous disease treatment. **Method:** Between May and July 2002, 567 GP have recruited 1049 female patients spontaneously consulting for CVD. The patients could fill in a series of validated questionnaires in order to evaluate the consequences of their disease. A patients subgroup with RA (treated with _ruscus aculeatus, hespéridine méthyl chalcone HMC & acide ascorbique Vit.C) prescription has been identified. **Objective:** To describe the impact in real conditions of treatment_s compliance. **Results:** The group with two tablets a day (n=135) is called the « non observant group : NOG », the group treated with the recommended dosage (4 tablets) a day is called _observant group: OG (n=831). Before treatment, both groups are comparable in terms of average age (44.1 vs 45), height and weight (BMI : 24.3vs 24.2). In order to strengthen the comparability of both groups, the risk factors have been compared: sedentary lifestyle, family history, underfloor heating, pregnancy. None of them is significant except sedentary lifestyle (NOG 55% vs 0:66%, p<0.0001, test ki2). At inclusion, specific and non specific quality of life (QoL) as well as daytime sleepiness were evaluated (CIVIQ, SF12 and Epworth). No significant difference has been observed between the NOG and the OG: CIVIQ : 34.3 v 32 , SF12 : Physical dimension: 48.2 v 46.2, Mental Dimension: 42.5 v 45, Epworth:7.2 v 7.8 . After a 7 day treatment, the same scales were administered, in the NOG, no QoL scale has improved. In the OG, SF-12 Mental dimension, CIVIQ and Epworth scores significantly improved at D7 versus D0 (with p respectively <0.001, =0.01, <0.001) **Conclusion:** The compliance with treatment at recommended dosage clearly shows an improvement of the specific and non specific quality of life scales at 7 days.

Abstract 189

CHRONIC VENOUS DISEASE: DISEASE AGE IMPACT

Charles Taieb, MD, Public Health , Health Economics & Quality of life, I.R.P.F., Boulogne Billancourt, France, Jean Jérôme Guex, Pr, Société Française de Phlébologie, Nice, France, Nathalie Marionneau, PhD, Public Health , Health Economics & Quality of life, I.R.P.F., Boulogne Billancourt, France, Robert Launois, Pr, REES, Réseau d'Evaluation en Economie de la Santé, Paris, France, Eric Myon, PhD, Public Health , Health Economics & Quality of life, I.R.P.F., Boulogne Billancourt, France

Introduction: The disease age impact has often been studied for pathologies such as the diabetes; paradoxically, it has been very little studied for chronic venous disease treatment. Method: Between May and July 2003, 567 GP have recruited 1049 female patients spontaneously consulting for CVD. The patients could fill in a series of validated questionnaires in order to evaluate the consequences of their disease. Objective: Describing the (P<0.001 test ki2) 4% of the patients of the 2 groups were offered « another prescription , most often doppler. We looked at the group of patients whose care has started since less than one month: life hygiene recommendations are given to 89%, a surgical opinion request is considered respectively for 1%, a sclerotherapy is proposed to 4%, finally an elastic contention prescription is found for 19% of the patients. None of these care differs from the less than 12 months group.(test ki2) Conclusion The initial care modalities of patients whose diagnosis age is included within the last 12 months is not taking into consideration all the possible therapeutic resources. Through healthcare professionals' education, care management could be improved.

Notes



***ISOQOL is dedicated to the promotion of excellence
in the science of health-related quality of life***

***THANK YOU to the following companies that
have contributed to this meeting:***

Benefactor

Pfizer, Inc.

Sustaining

**Amgen
Aventis**

Contributing

**AstraZeneca
Merck
Roche**

Supportive

**MEDTAP International, Inc.
Abbott Labs, Inc.**

Regular

**Health Assessment Lab
QualityMetric, Inc.**