

Abstract 1730**SO WHAT HAPPENED TO ALL THE PATIENTS? - EVENT CHARTS FOR SUMMARIZING CLINICALLY SIGNIFICANT CHANGES IN ONCOLOGY QUALITY OF LIFE DATA**

Pamela J. Atherton, Paul J. Novotny, Kate A. Clement-Brown, Jeff A. Sloan, Department of Cancer Center Statistics, Mayo Clinic, Rochester, MN

Purpose of Research: Event charts are a useful method of summarizing toxicity, disease progression, and survival in oncology clinical trials. In this poster we explore the use of event charts to summarize clinically significant changes in oncology patient quality of life (QOL) data. **Methods/Results:** Recent developments in QOL research have led to a number of definitions for clinically significant changes in oncology QOL measures. The approach of Osoba (JCO, 2000) and others suggest that on a 0-100 point scale, changes of <10 are small, 10-20 are moderate, and more than 20 points are large differences in QOL scores. This taxonomy is analogous to the tracking methodology invoked for the monitoring of toxicity data via National Cancer Institute Common Toxicity Criteria guidelines. We used this categorization and event chart methodology to summarize QOL data over time for patients enrolled in oncology clinical trials. Data were drawn from North Central Cancer Treatment Group (NCCTG) clinical trials. We compared event chart data to standard Kaplan-Meier time-to-event graphical representations and summary statistics. The event chart method proved superior to plotting raw scores over time since at-risk individuals are either identified with greater facility and censored or missing data are incorporated more readily due to the intent-to-treat nature of the method. Furthermore, event charts identified time points where patients may have experienced potential crises in QOL and where interventions may have been employed. **Conclusions:** The event chart provides an alternative method for summarizing individual patient QOL data over time. This methodology has the potential for use as a tracking device in oncology clinical trials. In this sense, we can record QOL "events" and potentially intervene based on the observed magnitude of change in scores.

Abstract 1133**ANALYSIS OF LONGITUDINAL QUALITY OF LIFE DATA WITH INCOMPLETE AND POSSIBLY BIASED FOLLOW-UP**

Michele Melia, Claudia S. Moy, Scott L. Zeger, Department of Ophthalmology, Johns Hopkins University, Baltimore, MD

The Collaborative Ocular Melanoma Study (COMS) is a multicenter randomized trial sponsored by the National Eye Institute and National Cancer Institute. Its primary goal is to determine whether patients treated by enucleation (removal of the eye) versus radiation (eye is preserved) have better overall survival, cancer free survival, and visual acuity. Annual health (HR-QOL) and visual functioning (VF-QOL) assessments using the SF-36, ADVS, NEI-VFQ, and HADS were added as outcomes while patient accrual was already in progress. As of December 1999, of 1317 patients enrolled in the trial, 600 patients had received assessments only after randomization, 209 had assessments both prior to and after randomization, and 508 had never had an assessment. Reasons for no assessment included death prior to implementation of assessments, refusal, and inability to complete an assessment due to hearing or other communication difficulties, or having a health condition, such as advanced Alzheimer's disease, that precluded assessment. Statistical analyses of these data present many challenges. Follow-up data for COMS patients consist of annual interviews for up to 4 years, with complete, partial, or no overlap between patients in the post-randomization time period in which assessments were made. Furthermore, missing assessments probably are not independent of HR-QOL or VF-QOL outcomes. Analytic strategies to account for these problems include use of linear mixed models that adjust for correlation in an individual's scores and allow for variation in scores across individuals, and development of a probability model for participation in the HR-QOL and VF-QOL assessments using other available trial data on health and vision status, that will be used to weight observed data to adjust for non-response. Our experience with these strategies will be discussed and methods will be illustrated using selected scales from the interview instruments.

Abstract 1106**MISSING DATA: CHALLENGES IN THE ANALYSIS OF THE BREAST EXPRESSIVE-SUPPORTIVE THERAPY (BEST) TRIAL.**

Louise Bordeleau, Andrea Bezjak, John Paul Szalai, Carol A. Sawka, Pamela J. Goodwin, Department of Medical Oncology, University of Toronto, Toronto, ON, Canada

The BEST study is a randomized trial of group psychosocial support versus standard medical care in women with metastatic breast cancer (MBC). The study objectives were to determine the effect of weekly support groups on duration of survival (primary objective), psychosocial functioning and quality of life (QOL) (secondary objectives). Eligible patients (n=235) completed the EORTC QLQ-C30 questionnaire in addition to a number of validated psychosocial questionnaires. Questionnaires were administered at baseline and every four months for a total of one year. Accrual was completed in December 1997. Patients with MBC often experience disease and treatment related morbidity and mortality. This inevitably leads to missing assessments as a consequence of patient non-compliance with the potential to reduce the power of the study and introduce bias in the final results. In the BEST study, preliminary analyses have shown an overall questionnaire completion rate of approximately 70% (adjusting for terminal illness). Missingness of data was mostly related to missing forms as opposed to missing items. Data from the EORTC QLQ-C30 questionnaire will be summarized and methodologic issues explored. Missing QOL data will be presented in various formats as previously suggested (Machin D et al. Stat Med. 1998;17:711-724). Type of missing data (missing forms, missing items) will be further categorized. In addition, the process(es) that caused missing data (missing completely at random, missing at random, missing not at random) will be identified. Based on the nature of missing data, potential types of analyses to evaluate the outcome of the BEST study will be explored. Given the lack of standard methodology for handling missing data in longitudinal QOL studies, we intend to analyze the data using different missing data models to compare outcome and establish the sensitivity of our results to the method of analysis.

Abstract 1247**MULTIPLE IMPUTATION METHOD FOR EORTC QLQ-C30 WITH MISSING DATA IN CANCER CLINICAL TRIALS**

Benny C. Zee, Ricky Mak, Dongsheng Tu, Clinical Trials Group, NCIC, Queen's University, Kingston, ON, Canada

In cancer clinical trials, health related quality of life (QOL) is usually measured by a patient self-administered questionnaire with multi-dimensional domains. It is a common problem to have missing observations in the form of missing items within a domain or missing for the whole questionnaire. Methods commonly used to analyse data with missing values include: 1) complete-case method; 2) "subscale mean" - impute a mean value of the domain; 3) "subscale mean 50%" - impute subscale mean when we have at least 50% of items in the domain; 4) "items mean" - impute the average of the items for those subjects who have answered that question; Other methods include: 5) "single imputation" - impute a value which is randomly selected from the corresponding item in the set of similar subjects; 6) "matched items mean" - impute the mean of the corresponding items instead of a randomly selected value in the set of similar subjects; 7) "multiple imputation" - impute the average of a set of candidates generated using the selected set of similar subjects as in single imputation. The process of creating the set of similar subjects and the generation of the candidate list reflects the sampling variability of the missing items. Patients who have complete QOL data from two antiemetic cancer clinical trials were used to evaluate the above methods. Missing QOL data were generated by two approaches: 1) missing at random; 2) the probability of missing data is a function of the subjects' global QOL score so that patients with poor global QOL are more likely to be associated with higher number of non-responses. The non-response rates of 10%, 20%, 30% were used in the simulation. All methods were compared on the basis of accuracy with respect to the sum of squared deviation of the predicted total score from the imputed data set versus the observed total score from the real complete data set. The results show that "subscale mean 50%" has reasonable performance, and "multiple imputation" method out performs other methods.

A NEW APPROACH TO ANALYSIS OF LONGITUDINAL INCOMPLETE QL DATA USING IDENTIFYING RESTRICTIONS FOR PATTERN MIXTURE MODELS.

Desmond Curran, Geert Molenberghs, Herbert Thijs, Geert Verbeke, Department of Biostatistics, ICON Clinical Research, Dublin, Dublin, Ireland

Missing data mechanisms such as missing completely at random (MCAR), missing at random (MAR), and missing not at random (MNAR) (Little and Rubin 1976) are usually presented in the selection modelling framework ($S: f(y, d) = f(y) f(d/y)$). Recently, Little (1993) has suggested pattern-mixture models ($PM: f(y, d) = f(d) f(y/d)$) as a valuable alternative to selection models. Although S and PM models are interchangeable from a probabilistic point of view, in practice they encourage different kinds of simplifying assumptions. One attraction of S models is that they fit naturally into Little and Rubin's taxonomy, whereas for PM models this is less clear. Molenberghs, Michiels, Kenward and Diggle (1998) showed that using Available Case Missing Values identifying restrictions for PM is equivalent with MAR. This has been used by several authors to contrast both modelling families. In order to compare S and PM it is necessary to develop strategies to deal with PM which have similar features to those used in S. We present strategies to fit PM models using several identifying restrictions: complete case missing value (CCMV), available case missing value (ACMV) and nearest case missing value (NCMV). We show how these can be extended in the form of a sensitivity analysis. These strategies are applied to QL data collected in an EORTC phase III trial: a prospective multicenter randomized study comparing flutamide versus prednisone in hormone resistant metastatic prostate cancer. Using identifying restrictions several values are imputed taking into account all relevant information. The so-completed data is then analyzed and inferences drawn using multiple imputation technology. All steps have been incorporated into a series of SAS macros.

Abstract 1257

DO DOCTORS KNOW WHAT ISSUES ARE IMPORTANT TO CHILDREN WITH INFLAMMATORY BOWEL DISEASE?

Hester J. Loonen, Bert H. Derkx, Adrian G. Thomas, Anne M. Griffiths, Department of Paediatric Gastroenterology, Emma Children's Hospital, Academic Medical Center, Amsterdam, The Netherlands

Objective: To analyse the concordance between items patients report to be important in their lives with IBD, and those items doctors think are important to their patients with IBD. Methods: An item reduction questionnaire consisting of 96 items previously identified by a large focus group (n=81) was administered to paediatric gastroenterologists specifically familiar with treating paediatric patients with IBD. Subjects: 21 paediatric gastroenterologists and 117 IBD patients aged 10-18 years. Results: Only one item that children considered in the top 10 of important items was also in the top 10 physicians thought would be important to them (abdominal pain and cramps). Important features are the physically oriented items physicians considered important as opposed to the more socially and emotionally oriented items children considered important. Most striking is the item: having to take medication which was scored number one by the children and only number 74 by physicians. Conclusion: Physicians' ideas of what issues and concerns are most important to children with IBD differ substantially to those stated by the children themselves. This study strongly emphasises the importance of including multiple sources of information in the item-gathering phase of developing a quality of life questionnaire. Furthermore it taps on the need for including a psychosocial approach in addition to traditional physical treatment in order to minimise disease burden and health related problems in children with IBD.

Top 3-ranking of item-importance for physicians and children

Item-importance ranking	By children	By physician
1	Having to take medication	Having stomach aches and cramps
2	Worries about having a flare-up	Bothered by stomach aches and cr
3	Worries about life-long disease	Having diarrhoea or frequent sto

THE PROXY PROBLEM: PARENT REPORT VERSUS SELF-REPORT OF THE CHILD HEALTH QUESTIONNAIRE

Hein Raat, Marie-Louise Essink-Bot, Jeanne M. Landgraf, Reinoud J. Gemke, Gouke J. Bonsel, Department of Public Health, Erasmus University, Rotterdam, The Netherlands

Outcome studies in pediatrics often comprise a wide age range, e.g. toddlers to (young) adults. Most QOL-instruments cannot cover this age span and a set of instruments has to be applied: e.g. for pre-school-, elementary school- and adolescent age. Moreover proxy measures have to be used for young children and self-report measures for adolescents. This study evaluates the agreement between parent- and child health reports as measured by the Child Health Questionnaire-Parent Form (CHQ-PF 5-16 years) and -Child Form (CHQ-CF 10-18 years). In this study 153 Dutch schoolchildren (10-12 years) and their parents were invited to complete the CHQ-CF and -PF. The response of children/parents was 97/87% resulting in 120 parent/child pairs. The PF-CF Intra Class Correlation-coefficients were rather low: Physical functioning and two Role functioning scales showed no significant ICC; 7 other scales showed ICCs from .25 - .34. E.g., for reasons of comparison, the self-report inventory of chronic diseases showed a parent-child ICC of .53. On the group level 4 out of 10 scale-means had no significant parent-child differences. Children reported more problems on average regarding 4 scales with a physical or role functioning content. Parents more often reported behavioral- and family problems. Meanwhile, in this study, both CHQ-PF and -CF showed very good internal consistency of the scales, adequate test-retest characteristics and acceptable construct validity. We conclude that the CHQ-PF and -CF are both valid and reliable instruments, but measure (partly) distinct perspectives on child health. Preferably both should be applied if possible.

Abstract 1311

HRQOL AS AN OUTCOME ASSESSMENT PARAMETER IN DISEASE SPECIFIC OUT-PATIENT REHABILITATION OF CHILDREN AND ADOLESCENTS

Ulrike Ravens-Sieberer, Michael Redegeld, Monika Bullinger, Department of Medical Psychology, University of Hamburg, Hamburg, Germany

The assessment of HrQoL of patients is increasingly discussed in rehabilitation research. The majority of studies however address adults. In rehabilitation medicine studies about perception of well-being and function with chronic diseases before and after in- or outpatient rehabilitation from the children's and the parent's perspective are still rare. The longitudinal multicenter study describes the results of QoL assessment of children and parents at start, at end, after 3 and 12 months of out-patient treatment. 1099 children (53% female, age 8-18, 613 with obesity, 486 with asthma/atopic dermatitis) in 7 rehabilitation clinics and their parents filled in questionnaires relating to HrQoL outcomes (e.g. the German generic core Kindl with newly developed disease specific modules for asthma, atopic dermatitis and obesity) as well as clinical and psychosocial variables. Data analysis included psychometric testing as well as correlation and regression analysis, analysis of variance and analysis of change over time. Psychometric testing of the generic and disease-specific QoL-questionnaires showed good reliability and validity for chronically ill children and for parents (alpha up to .89, high correlation with other QoL measures). Predictors for high QoL in children were self esteem and locus of control, while clinical indicators (such as disease severity or duration) were of less importance. Intercorrelations between children's and parent's quality of life assessment were moderate (approx. r = .40). Change over time data co-varied with the perceived quality of care, effect size estimation (Kazis) showed treatment effects on QoL. Results so far indicate that QoL can be assessed reliably and validly with the Kindl and it's new modules in ill children. Inclusion of children's self report data in outcome assessment in rehabilitation research is feasible and should be performed. Disease specific results also affect training programs for children and parents.

Abstract 1381**VSP-AS: TIME FOR A PRAGMATIC STEP**

C SAPIN, MC SIMEONI, A CLEMENT, F PELLEING, JL SAN MARCO, P AUQUIER, PUBLIC HEALTH DEPARTMENT, MARSEILLE, FRANCE

The VSP-As is a French, valid, reliable and sensitive Health-Related Quality of Life (HRQL) 42-item measure for adolescents attending school (aged 10-17), based on their viewpoint. It produces 7 dimensions: Psychological Well-Being, Energy, Friends, Parents, Leisure, School, Sentimental and Sexual Life and an index; it was already completed by 6,000 adolescents. The survey aimed at showing the contribution of HRQL measures to systematic checkups in general population, especially in school checkups. The target population was the pupils in 8th grade in the high school of our south-eastern county. A stratified sampling was used (sampling fraction=1/10), taking into account 2 stratifying variables: location of school in a socio-economically disadvantaged area (ZEP) or not, in a rural (urban) area. During the school checkup adolescents fulfilled the VSP-As while physicians filled a standardized clinical form: sociodemographic data, adolescent's health status assessment (previous history, vision, hearing, acute or chronic disease, Clinical Global Impression), information about adolescent's way of life (sport activity, smoking, contraception), Visual Analogue Scales measuring domains related to VSP-As dimensions. In this study, 2,000 pupils were included: 3% in rural ZEP, 4% in rural not ZEP, 23% in urban ZEP, 70% in urban not ZEP. The first results showed that adolescents' self-reported VSP-As scores were consistent with objective clinical criteria: the more impaired the health status was, the lower the dimension scores were. However, correlation levels between VSP-As scores and classical usually used indicators of risky behaviors were low; discrepancies were found between the subjective health assessment provided by physicians and the subjective adolescents' VSP-As evaluation, especially concerning psychological and social fields. These results led us to propose such a measure for use in general population to point out vulnerable populations for which health prevention and care strategies could be planned. Granted by the Conseil Régional PACA

Abstract 1469**HEALTH STATUS OF OVERWEIGHT/OBESE AND UNDERWEIGHT CHILDREN: A POPULATION BASED SURVEY**

Melissa Wake, Louisa A. Salmon, Elizabeth B. Waters, Centre for Community Child Health, RCH, Parkville, Victoria, Australia

Objective: Childhood overweight/obesity is associated with poor physical and psychosocial health in clinical samples. Little is known about health effects of underweight. We examined parent-reported child health and extremes of BMI in a population sample of primary school children. **Methods:** Sample: a stratified two-stage random cluster sample of 24 primary schools representative of Victoria, Australia. Measures: BMI (wt/ht²); the Child Health Questionnaire (CHQ), a 13 scale 50-item parent-completed measure of health status. **Results:** Data were available for 2863 children aged 5-13 years (50.5 % male). Ranked Self-Esteem and General Health scores were lower for obese boys, General Health scores were lower for obese girls and Self-Esteem scores lower for over and underweight girls (Kruskall Wallis). On logistic regression analyses obese boys were more likely to receive low scores on 7 of 12 scales: Physical Functioning (OR 2.8), Bodily Pain (OR 1.8), General Health (OR 3.5), Mental Health (OR 2.8), Self-Esteem (OR 1.8), Parent Impact-Emotional (OR 1.7) and Parent Impact-Time (OR 1.9). Obese girls were more likely to receive low scores on only 2 of 12 scales: General Health (OR 2.1) and Self-Esteem (OR 1.8). Underweight boys were more likely to receive low scores on Bodily Pain (OR 2.1), and underweight girls on Physical Functioning (OR 1.9), Role-Physical (OR 1.9), General Health (OR 2.2), and Parent Impact-Emotional (OR 1.8) scales. **Conclusions:** Parents reported health deficits over a wider range of health domains for overweight/obese boys than for girls. Most overweight/obese and underweight boys and girls did not receive low health scores.

Abstract 1620**LOWER SCORES ON THE HOSPITAL ANXIETY & DEPRESSION (HAD) SCALE FOR A POPULATION-BASED SAMPLE OF BREAST CANCER PATIENTS COMPARED WITH POPULATION CONTROLS: PSYCHOMETRIC ARTEFACT OR REAL DIFFERENCE?**

Richard H. Osborne, Gerald R. Elsworth, Frans J. Oort, Mirjam A. Sprangers, John L. Hopper, Centre for Genetic Epidemiology, University of Melbourne, Melbourne, Vic, Australia

The HAD is a widely used 14-item 2-scale instrument that has undergone several psychometric studies. While the psychometric properties of the scale are generally strong its qualities vary according to the test sample. We administered the HAD to population-based samples of 728 women with breast cancer (BC) and 158 reference women. Contrary to expectations we found that the BC group had lower anxiety & lower depression scores than the reference group, although the standardised mean differences (ie effect size) were small: ES=-0.17 (p=0.06) for anxiety and ES=-0.26 (p<0.01) for depression. After adjusting for group differences (age) and potential cofounders (family history, marital status, education) using multiple logistic regression, significant differences remained. Do these findings reflect true differences or are they caused by item bias? An item is said to be biased when varying item scores cannot be sufficiently explained by varying levels of the trait that the item supposedly measures. We used structural equation modeling to detect bias in the HAD items. Analyses suggested that some bias was present in 1 Anxiety and in 2 Depression items. The direction of the bias in the Anxiety item was against the reference group (ie reducing group mean score) whereas the bias in the Depression items was against the BC group. Recalculation of the raw scale scores without the biased items increased the difference on the Anxiety scale (ES=-0.20 p=0.03) & decreased the difference on the Depression scale (ES=0.12 p=0.16). Still, the respective changes of effect sizes are not significant (alpha=0.05). In summary, we found some evidence of measurement artefact but it cannot explain the unexpected lower anxiety scores of the BC group. However, removal of biased items from the Depression scale diminished the between group difference found earlier.

Abstract 1626**QUALITY OF LIFE IN RELATION TO SURGERY, RADIATION AND ADJUVANT SYSTEMIC TREATMENT IN WOMEN WITH BREAST CANCER**

Rohini W. Hawaldar, Rajendra A. Badwe, Aley Yamaa, Srinivas Rayabhattachanwar, Clinical Research Secreariat, Tata Memorial Hospital, Mumbai, Maharashtra, India

The European Organization for Research and Treatment of Cancer (EORTC) module QLQ-C30 was used to assess the quality of life (QOL) in terms of physical health, functional status, psychological distress and social functioning in Indian women with operable breast cancer(OBC). The breast cancer specific module, BR-23 was used to assess disease/treatment related symptoms. The modules were translated according to EORTC translation procedure into Hindi, Marathi & Gujarathi. Two hundred sixty two patients treated for OBC were assessed for their QOL at 3 different time points, the first, on completion of surgery, the second, during adjuvant systemic therapy and the third, on completion of adjuvant treatment. The average time for completing the questionnaire was 10 min. Reliability and validity of the questionnaire was tested by Cronbach's alpha (0.68-0.95) and item-scale correlation (0.65-0.94). Physical, emotional and social function of a patient did not perceptibly change over a period of 6 months. Day to day activity of the patient, cognitive function and global QOL significantly deteriorated from visit 1 to 2 but improved by the time patient completed adjuvant therapy. Body image was grossly superior with breast conservation (BCT) as compared to mastectomy at the first visit (p < 0.00001). This perception reduced at the second visit but third visit BCT is superior to mastectomy (p=0.03). Physical, emotional and role functioning scales showed adverse effect of chemotherapy and radiotherapy but not of hormone therapy. Radiotherapy (p=0.029) and chemotherapy (p=0.002) significantly affected Global QOL whereas type of surgery and hormone therapy did not have any impact on global QOL. QLQ C-30, BR-23 can be used to assess QOL in Indian patients. Chemo-radiation adversely affected QOL and should be used when clear evidence of benefit exists. BCT maintained satisfactory body image in patients with OBC and should be offered wherever feasible.

Abstract 1759

HEALTH-RELATED QUALITY OF LIFE (HRQL) CORRELATES OF PERCEIVED OVERALL HEALTH IN SPANISH- AND ENGLISH-SPEAKING BREAST CANCER POPULATIONS

Carol M. Moinpour, Gary W. Donaldson, Geraldine V. Padilla, Patricia A. Ganz, Sandra L. Pineros, Laura C. Lovato, Department of Southwest Oncology Group Statistical Center, Fred Hutchinson Cancer Research Center/MP557, Seattle, WA

Purpose. The purpose of this analysis was to examine which components of a comprehensive HRQL assessment determine patients' rating of overall health. **Methods.** Patients with breast cancer (early and late stage disease) completed questionnaires in English and Spanish that address the following HRQL dimensions: breast-specific symptoms, general symptoms, physical, role, emotional, and social functioning. Multiple regression analysis was used in several stages to predict the overall health rating scale from the symptom and functioning measures of the Southwest Oncology Group (SWOG) HRQL questionnaire. Of interest is the magnitude of the drop in R² on eliminating the relevant subsets of predictors from the regression equation. **Results:** The full model without interactions adequately describes the determinants of the overall health rating. No coefficients differ significantly by language group. Adjusted simultaneously, disease-specific symptoms have no significant effect on overall health rating, while general cancer symptoms have a noteworthy association. Each of the four functioning measures also displays significant unique predictability, with role and social functioning explaining more variance uniquely than either physical or emotional functioning. Considered as a set, the four functioning measures are somewhat more influential than either the individual or combined symptom measures. **Conclusions.** Although general symptoms and all four domains of HRQL are probably important considerations when patients evaluate their overall health, the role of specific symptoms is redundant with the other measures. The importance of general HRQL measures in predicting perceived health does not vary by language group.

Abstract 1399

PROFILES OF LONGITUDINAL QOLSCORES AMONG PEOPLE WITH BREAST AND COLO-RECTAL CANCER

Barbara A. Elliott, Ron R. Regal, Colleen M. Renier, Thomas E. Elliott, Department of Family Medicine, UMD-School of Medicine, Duluth, MN

The purpose of this study was to describe the similar profiles of longitudinal QoL scores evident among people with stages 0, 1, 2, and 3 breast cancer and people with stages 0, 1, 2, and 3 colo-rectal cancer. In a study designed to enhance rural practitioners' cancer care, patient outcomes including QoL were measured. A population of 237 recently diagnosed breast and 103 colo-rectal cancer patients in 13 rural Midwestern (USA) communities were interviewed and invited to complete QoL (FLIC) forms nine times over 2 years: at baseline (within one month of diagnosis) and at 2, 4, 6, 9, 12, 15, 18, and 24 months. The resulting profiles were analyzed with mixed models repeated measures techniques. Statistically significant effects on QoL profiles for these patients included the length of time since diagnosis (p = 0.0001), type of treatment being experienced (chemo vs. no chemo: p = 0.01), type of cancer (p < 0.01), and the patients' scores on the Sense of Coherence scale (p = 0.0001). There were also statistically significant interactions between length of time and type of treatment (p < 0.0005), length of time and type of cancer (p < 0.05), and length of time and age of patient (p < 0.05). The profiles indicate that over time, the strongest influence on reported QoL is the type of treatment being experienced. The presentation will detail how QoL profiles change with time since diagnosis, treatment, cancer type and age.

Abstract 1486

POTENTIALLY MODIFIABLE PSYCHOSOCIAL AND TREATMENT RELATED FACTORS INFLUENCE QUALITY OF LIFE IN DIFFERENTIATED THYROID CANCER SURVIVORS

Luke K. Tan, Lincoln Tan, Arul Ernest, Felix Sundram, Julian Thumboo, Dept. of Otolaryngology, National University of Singapore, Singapore, Singapore

Background: As subjects with differentiated thyroid cancer have survival rates exceeding 80% after 20 years, their quality of life (QoL) is an important issue. Although QoL of these subjects is somewhat lower than that of the general population, little is known about factors influencing QoL in such subjects. **Aim:** To identify factors influencing QoL in survivors with differentiated thyroid cancer. **Methods:** 250 subjects with differentiated thyroid cancer were assessed using identical English or Chinese postal questionnaires containing the Short Form 36 Health Survey (SF-36) and assessing demographic, psychosocial, disease and treatment related variables. Separate linear regression models were constructed using each SF-36 scale as an outcome variable, and including these variables and questionnaire language as predictor variables. **Results:** 150 subjects (Chinese: Malay: Indian 129:12:9) returned questionnaires after 3 mailings. Their median age of 46.9 years; 28% were males, 54% were working, 11% had ceased thyroxine within 6 weeks of the survey and all had previous thyroid surgery. Psychosocial factors influenced scores for all SF-36 scales, with higher levels of learned helplessness consistently associated with lower QoL, and better family support with improved QoL. Thyroxine cessation was associated with poorer social functioning and mental health and better physical functioning, and working with better physical and emotional functioning. Age, gender, socio-economic status, duration from diagnosis and questionnaire language also influenced scores for several SF-36 scales. **Conclusion:** QoL in surviving subjects with differentiated thyroid cancer is influenced by demographic, socio-economic and potentially modifiable psychosocial and treatment related factors.

Abstract 1726

MINIMALLY PERCEIVED DIFFERENCE OF THE KING'S HEALTH QUESTIONNAIRE IN PATIENTS WITH SYMPTOMS OF OVERACTIVE BLADDER: A COMPARISON OF RESULTS FROM TWO STUDIES

Pat R. Reese, Gary J. Okano, Andreas M. Pleil, Con J. Kelleher, Strategic Outcomes Services, Inc., Research Triangle Park, NC

The purpose was to compare estimates of minimally perceived difference in the King's Health Questionnaire (KHQ) for patients with overactive bladder (OAB) using data from two studies and a method similar to Juniper et al. Data from two multinational RTCs comparing the efficacy and safety of two dosage forms of tolterodine versus placebo (n = 1,529) and tolterodine versus oxybutynin (n = 827) for treatment of OAB were used without consideration for treatment or translation. Study 1 used a global question of perceived bladder condition at baseline and after 12 weeks of treatment with a 6-point scale (no problems - many severe problems). Study 2 used a global rating of treatment benefit after 6 months of treatment with a 3-point scale (no benefit - much benefit). The difference in mean KHQ scores for patients indicating no change in bladder condition or treatment benefit and those reporting a 1-point improvement or small treatment benefit was considered the minimally perceived difference. Minimally perceived differences were smallest for Personal Relationships and Symptom Severity measures and most variable between the two estimates for Emotions and Sleep and Energy.

Difference in Mean KHQ Scores of Patients Reporting Improvement (Study 1) or Treatment Benefit (Study 2) and Patients Reporting No Change

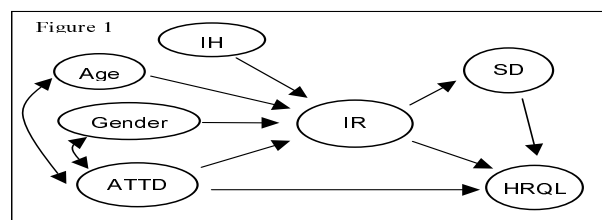
	Study 1 (n=405)	Study 2 (n=225)
Incontinence Impact	-7.91	-5.62
Role Limitations	-6.29	-7.48
Physical Limitations	-7.67	-5.02
Social Limitations	-4.44	-2.08
Personal Relationships	-2.12	-1.41
Emotions	-5.15	-9.13
Sleep and Energy	-4.91	-9.60
Severity (coping) Measures	-4.49	-3.65
General Health Perception	-3.05	-5.23
Symptom Severity	-1.68	-1.47

In conclusion, further evaluation is needed to access the impact of different criterion variables and different countries on the estimate of minimally perceived difference in the KHQ.

TOWARD A THEORY OF TECHNOLOGY DEPENDENCY AND HEALTH-RELATED QUALITY OF LIFE

Susan F. Marden, University of Maryland, Silver Spring, MD

Technology dependency is the reliance on health care technology to resolve a health-related problem.¹ This paper proposes a new mid-range theory characterizing the relationship between technology dependency and health-related quality of life (HRQL). Generated through the synthesis of 3 theoretical perspectives 1,2,3 and empiric research 4,5, the theory postulates that attitudes toward technology dependency (ATTD) affect HRQL (see Figure 1). It also suggests that the ATTD-HRQL relationship is influenced by a person's illness representation (IR; common sense beliefs regarding illness), symptom distress (SD), illness history (IH), age, and gender (see Figure 1). The theory, currently undergoing empirical testing in a sample of implantable defibrillator recipients, is intended to serve as a framework for clinical research on the impact of a wide variety of technological interventions on HRQL.



¹Sandelowski, M. (1993) The Conceptual Framework of Technology Dependency in Health Care

²Spilker, B. (1996). Relationship of Clinical, Safety, and Efficacy Data to Quality of Life: The Filtering Phenomenon

³Leventhal et al (1984). Self-Regulation Model of Illness Cognition and Behavior

⁴Grady et al. (1995) Predictors of quality of life in patients with advanced heart failure awaiting transplantation. *The Journal of Heart and Lung Transplantation*, 14, 2-10.

⁵Heijman, M. (1997). Coping and adaptive outcome in chronic fatigue syndrome. *Journal of Psychosomatic Research*, 45, 39-51.

Abstract 1310

THE USEFULNESS OF CRISIS THEORY IN EXPLAINING EXPERIENCED LEVELS OF QUALITY OF LIFE

Mirjam A. Sprangers, Hanneke C. De Haes, Department of Medical Psychology, Academic Medical Center, Amsterdam, Netherlands

Based on the premises of crisis theory, we expected cancer patients in crisis to report a poorer quality of life (QL) and cancer patients post-crisis to report a similar level of overall QL in comparison to healthy individuals. To explain these hypothesized findings, we expected the coping resources of patients in crisis to be equally and those of patients post-crisis to be more effective as compared to those of healthy individuals. The sample comprised: (a) 217 consecutive cancer patients in the acute phases of their illness (patients in crisis); (b) 192 disease-free cancer patients (patients post-crisis); and (c) 201 randomly selected healthy individuals. Mailed QL measures included overall QL, physical and psychological distress and role activity. Coping resources were assessed with questionnaires measuring self-esteem, neuroticism, and coping behavior. As expected, patients in crisis reported lower levels of overall QL and role activity and higher levels of physical and psychological distress in comparison to healthy individuals. Patients post-crisis reported a similar level of overall QL and psychological distress and a higher level of physical distress and a lower level of role activities than healthy persons ($p < .01$). There were no significant differences between the mean levels of coping resources between the respective groups. Two-way analysis of variance indicated a group X coping resource interaction-effect on overall QL for self-esteem ($p < .01$). As expected, the amount of variance of overall QL explained by self-esteem was largest for patients post-crisis (27%) and comparable for patients in crisis and healthy individuals (10% and 11%). Patients in crisis reported a poorer QL and patients post-crisis a similar overall QL as compared to healthy individuals. Patients in crisis were not able to make their coping resources more effective, whereas patients post-crisis succeeded in enhancing the effectiveness of self-esteem in restoring their QL as compared to healthy persons.

Abstract 1816

CHAOS THEORY AND QUALITY OF LIFE (QOL): A NEW PARADIGM?

Rao S. Pippalla, School of Pharmacy, Howard University, Washington, DC

The main purpose of the study was to explore applications of Chaos Theory, Complexity Science and Non-linear models to the studies of QOL. Quality Of Life is "a complex construct." and based on earlier definition of health by WHO. But WHO recently proposed a new definition for health as "a dynamic state of complete physical, mental, spiritual and social well-being and not merely the absence of disease or infirmity." (52nd WHO Assembly). Now, there is a compelling need to revisit our definitions of QOL and reevaluate its assessments embracing dynamic non-linear models. "Intellectual comfort", simplicity, and non-availability of high-speed computers might have contributed to our heavy reliance on linear models in QOL research. Because it was observed that "non-linearity is health and linearity is death." I am hypothesizing that every assessment of QOL at any point in time (cross-sectional/longitudinal) is a measure of Complex Adaptive System (CAS). As per Complexity Science 'whole' impacts 'parts' and 'parts' modify 'whole' constantly and there exists a constant feed back among them. In Complexity Theory 'humans' who answer questions are clustered (Kauffman's root square analysis) than questions are grouped (factor analysis). Hence eight data sets ($n=297$) on different diseases were revisited for the purpose. Besides, the main tenet of Chaos theory is its 'sensitive dependence on initial conditions' (Butterfly effect) and thus brings strength to existing models in predicting disease progression with corresponding changes of QOL. As one of the themes of the conference is "measurement of meaningful change in HRQOL" two data sets (hypertension, $n=87$; discharge counseling project, $n=23$) were reanalyzed to compare and contrast non-linear Phase-Plane models with traditional paired t-tests. Lastly, Complexity economics identifies pharmaceuticals as the products of 'increasing returns' than traditional 'diminishing returns'. Its impact on Pharmacoeconomics including productivity measurements will also be addressed in the study.

Abstract 1132

TESTING MODELS OF QUALITY OF LIFE (QOL) II: EXPLORATORY AND CONFIRMATORY METHODS

Joseph R. Pellizzari, David R. Evans, London Health Sciences Centre, The University of Western Ontario, London, ON, CANADA

A generic QOL model (Evans, 1994) including personality characteristics (self-esteem, dispositional optimism, neuroticism, extraversion), subjective well-being (positive and negative affect, global and domain life satisfaction), and QOL outcomes (behavioural functioning, health-related QOL) was used as a guiding framework in order to test various hypothesized relationships. Pathology based measures (anxiety and depression) were also included for discriminant validity purposes. Accordingly, a self-report assessment battery was designed and administered to two adult, community samples of London, Ontario ($n=314$, $n=273$). Empirical testing proceeded via exploratory (exploratory factor analysis - EFA) and confirmatory (CFA) methods. Examination of zero-order correlations generally revealed moderate to high correlations both within and between the several classes of well-being/QOL variables. This was further reinforced by EFA and CFA. EFA revealed a 3 factor solution (49.5% of the variance) with Psychological Well-being (37.6%), Physical Well-being (7.3%), and Occupational Well-being (4.5%) as the three factors. Testing the hypothesized structural relationships among well-being/QOL components met with serious methodological challenge, namely, multi-collinearity. Relying on CFA, the hypothesized and nested measurement models solely within the previously identified Psychological Well-being component led to a 3 factor model: Personality/Affectivity, Life Satisfaction (cognitive-evaluative dimension), and Behavioural QOL. This correlated, 3-factor model offered an acceptable fit to the data ($CFI=.906$, $GFI=.850$, $AGFI=.811$). The existence of a higher-order factor was also considered to be likely. Attempts to cross-validate these findings were mildly successful. The results of these quantitative approaches to assessing the properties of the QOL construct have particular implications for the conceptualization and assessment of QOL and for clinical interventions across a wide variety of settings.

Abstract 1073

PATTERNS OF SELF-RATED HEALTH IN OLDER ADULTS BEFORE AND AFTER SENTINEL HEALTH EVENTS

Paula Dier, Donald L. Patrick

Purpose: To describe and compare patterns of change in self-rated health for older adults before Death, and before and after Stroke, Myocardial Infarction, Congestive Heart Failure, CABG or PTCA, hospital admission for Cancer, and Hip Fracture.

Subjects and Methods: 5,888 participants in the Cardiovascular Health Study (CHS), sampled from Medicare roles in 4 US communities, were followed up to 8 years. Mean age at baseline was 73. Self-rated health, including a category for death, was assessed at 6-month intervals, and all persons were followed for events. We estimated the percent who were healthy (Excellent, Very Good, or Good) each month in the five years before Death and in the two years before and after the other events, using graphs and multiple regression. We compared the patterns to a "no-event" group and to one another, using graphs and linear regression.

Summary of Results: For people who died, the percent healthy declined slowly until about nine months before death, when it dropped steeply. Health of people the same distance from death was unrelated to age, but men and whites had higher values. For the other events, health declined somewhat before the event, dropped steeply at the event, showed some recovery and then declined further after the event. About 65 to 80% of the subjects were healthy 2 years before their event, but only 35 to 65% were healthy two years afterwards. Health changed as approximately the logarithm of time from event. Patterns were similar though less extreme for the "no-event" group. Years of Healthy Life can be calculated as the area under these curves.

Conclusion: Self-rated health can track the trajectories of changes near serious health events. Possible explanations for the large changes include response shift. We have described the trajectories in detail. Further work is now needed to explain, predict, and possibly prevent such changes in health.

Abstract 1690

PHYSICAL HEALTH STATUS IS PREDICTIVE OF SURVIVAL IN HIV+ PATIENTS

R. Murri, C. Del Borgo, I. IZZI, M. Fantoni, P. Limongelli, G. Deiana, QUAVISC GROUP, Department of Infectious Diseases, Catholic University, Rome, Italy

Objective: To evaluate the predictive value of both physical and mental health on clinical outcomes and survival of HIV-infected patients in the new antiretrovirals (ART) era. **Methods:** Prospective, multicenter, cohort study on Italian HIV+ patients at any stage of disease. From October 1997 to May 1998, both outpatients and inpatients were consecutively enrolled in 7 Italian Clinical centers. Questionnaire administration was planned every 6 months. The MOS-HIV Health Survey was used as tool of the study. Two summary scores for physical health (PHS) and mental health (MHS) were obtained. **Results:** 809 patients participated to the study. Mean age: 35.6 years, females: 31.8%, intravenous drug users: 48.2%; 41.6% of people belonging to the CDC's group A and 32.6% group C. Median CD4+ cells count: 259/ml; people with undetectable HIV RNA: 31.9%. 89.9% patients were taking any ART, of which 56.1% a protease inhibitors (PIs)-containing regimen. After a median of 697 (25th-75th percentiles: 637-762) days, 40 deaths and 67 first AIDS-related events were observed. Considering both deaths and first AIDS-related events, 95 episodes were registered. After adjusting for age, HIV transmission modality, CD4 cells count, therapy, monthly income, variables significantly associated at Cox model, to a poor survival were a PHS below the median (OR 11.80, 95% CI 2.77-50.14; p=0.0008) and belonging to CDC's group C (OR 2.31; 95% CI 1.02-5.21; p=0.04). Similar results were obtained considering both deaths and AIDS-related events in people taking ART containing PIs for more than 6 months (having PHS below the median: OR 2.88; 95% CI 1.22-6.77; p=0.01 and belonging to the CDC's group C: OR 3.12; 95% CI 1.37-7.09; p=0.006) **Conclusions:** Physical health status is one of the most strong predictors of poor clinical outcomes in a cohort of HIV+ patients at any stage of HIV disease. Introduction of health status assessment could be very useful in predicting future events and in recognizing people at high-risk of poor outcomes.

Abstract 1176

CAN THE HEALTH UTILITIES INDEX MEASURE CHANGE?

Jacek A. Kopec, Susan Schultz, Vivek Goel, Ivan J. Williams, Department of Health Care and Epidemiology, University of British Columbia, Vancouver, BC, Canada

The purpose of the study was to investigate the ability of the Health Utilities Index (HUI) to detect changes in health status in a general population cohort. We used data from the National Population Health Survey conducted in Canada in 1994/5 and 1996/7 (N=12,065). Responsiveness of the HUI was compared to that of the Self-Rated Health (SRH) scale (excellent, very good, good, fair, poor). Standardized response means (SRM) and the sensitivity coefficient (SC) were used to quantify responsiveness. We expected deterioration in health status among persons who were diagnosed between the two surveys with a serious medical condition, became restricted in activities, or required hospitalization, and among the oldest members of the cohort. Cross-sectional data from the 1996/7 survey were used to classify chronic conditions into mild, moderate, and severe. Overall, HUI scores improved between the two surveys in all age-sex groups, except men 65 years of age and older. Improvement was seen primarily in the cognitive domain of health status and, to a lesser extent, the emotional domain. In contrast, SRH scores declined in all age-sex groups during the same period. As expected, the HUI showed deterioration in health status among persons who developed a severe condition (SRM=-0.171), were hospitalized (SRM=-0.073), or became restricted in activity (SRM=-0.424). The SRMs were generally smaller for the HUI, compared with the SRH. The HUI was slightly more effective than the SRH in distinguishing between persons developing severe conditions and those remaining disease-free (SC=0.078). This study demonstrates that the HUI is able to detect changes in health status in a large population cohort. However, a single-item SRH scale may be more sensitive to change. The study also suggests that the aspects of health people consider when rating their overall health may differ from the conceptual model underlying the HUI. Validity of the HUI scales measuring each health attribute requires further study.

Abstract 1706

HEALTH-RELATED QUALITY OF LIFE AS A PREDICTOR OF FUTURE PHYSICIAN VISITS AND HOSPITALIZATION

Jeffrey A. Johnson, Barbara Conner-Spady, Simon Pickard, Institute of Health Economics, Edmonton, AB, Canada

The purpose of this study was to assess whether health-related quality of life (HRQL) is predictive of future physician visits and hospitalization. Data were collected annually over 3 years by mail survey as part of a longitudinal population survey in the province of Alberta, Canada. Surveys were mailed to a randomly selected sample of the adult population in June 1997. The 1518 baseline respondents (43% response rate) were then resurveyed in 1998 (75% retention rate) and 1999 (70% retained). HRQL measures included the physical and mental component summary scores (PCS12 and MCS12, respectively) of the SF-12, as well as the VAS and TTO-based index scores from the EQ-5D. Separate multivariate logistic regression models were used to assess the ability of each of the baseline HRQL measures to predict self-reported physician visits and hospitalization over the subsequent two years, after controlling for age, gender, and self-reported comorbidity, including depressive symptoms. In the multivariate models, 10-point decrements in baseline PCS12 scores were significant predictors for both physician visits (odds ratio (OR) 0.72; 95%CI: 0.58-0.92) and hospitalizations (OR 0.57; 95%CI: 0.46-0.72) in 1998; MCS12 scores were not predictive of resource use. A 0.1 increment on the EQ-Index was with a reduced risk of hospitalizations in 1998 (OR 0.83; 95%CI: 0.75-0.92) and 1999 (OR 0.86; 95%CI: 0.78-0.95), but was not predictive of physician visits in either follow-up year. Ten-point increments in the EQ-VAS was predictive of both physician visits (OR 0.85; 0.73-0.97) and hospitalizations (OR 0.77; 0.68-0.89) in 1998, with a similar OR in 1999. After controlling for demographic determinants and self-reported diagnoses, self-reported HRQL was a significant predictor of health care resource utilization. Increments of 10% on the HRQL measures were associated with a 15% to 40% reduction in self-reported physician visits and hospitalizations.

Abstract 1756**LONGITUDINAL ASSESSMENT OF QUALITY OF LIFE IN ADVANCED CERVICAL CANCER**

Howard T. Thaler, Richard M. McQuellon, David H. Moore, James P. Donnelly, David F. Cella, Department of Epidemiology and Biostatistics, Memorial Sloan-Kettering Cancer Center, New York, NY

Purpose of Research: Health-related Quality-of-Life (QOL) data in cancer patients are often multidimensional and measured repeatedly over time. QOL may both correlate with clinical variables and contain independent information about disease course and therapeutic effect. We sought a comprehensive statistical model to assess QOL in a randomized clinical trial (RCT). **Subject Sample and Methods:** The Gynecologic Oncology Group conducted a multicenter RCT comparing two chemotherapy (CTx) regimens in advanced cervical cancer (GOG-169). QOL measures complemented clinical outcomes related to tumor control and survival duration. Patients self-administered the Functional Assessment of Cancer Therapy (FACT) during the course of the first four cycles of CTx (FACT-G with five subscales + items specific to cervical cancer). Advanced disease has short expected survival time, and patients may go off study or expire before completing CTx and concurrent QOL measures. Thus data are not missing at random. **Statistical analysis should assess:** 1) variation within and between patients or groups, in overall levels and time-trajectories of QOL measurements; 2) clinical and demographic covariates correlating with outcome; 3) global and subscale-specific QOL trends; and 4) patterns of missing data relating to disease course. We used graphical, univariate and latent variable data analyses. **Summary of Results:** Of 269 patients with QOL data, 51% completed the questionnaires all 4 intended times. Treatment arm (Tx) and patients' initial Performance Status (PS) and age potentially impact QOL and clinical outcome. Globally, QOL was stable or moderately improving among patients completing all 4 assessments and worsened, with curtailed survival, among patients completing fewer. PS, age and Tx correlated differentially with QOL and clinical outcome. PS was a highly significant univariate predictor of survival, but baseline FACT Physical Well-Being subsumed all its prognostic value.

Abstract 1570**SOAP-51: EVALUATION OF CLIENT-EXPRESSED PREFERENCE FORMATS FOR WEIGHTING QUESTIONS IN A SCHIZOPHRENIA-SPECIFIC QUALITY OF LIFE SURVEY INSTRUMENT**

Gerald E. Schumacher, Judith Barr, Jessica Goren, Susan Ohman, Annette Hanson, School of Pharmacy, Northeastern University, Boston, MA

Objective: To determine a method, using client preferences, to weight the 51 items of the Schizophrenia Outcomes Assessment Project (SOAP-51), our validated schizophrenia-specific quality of life instrument. **Methodology:** 30 individuals with schizophrenia ("clients"), giving informed consent, completed surveys and gave qualitative interviews at their local community psychiatric clinics. Each responded to 8 of the 51 items using 6 response formats: (A) horizontal, 10-cm line, intermediate numbers of 1-3 and qualifying words, (B) horizontal, 10-cm line, intermediate numbers of 0-100 and qualifying words, (C) verbal "feeling thermometer", (D) qualifying words only, no quantification, (E) horizontal, qualifying words-anchored, numberless, (F) vertical, 10-cm line, intermediate numbers of 0-100. Participants were interviewed to determine their favorite, 2nd favorite and least favorite scale formats. A chi-square analysis was used to determine if observed versus expected responses differed within consumers' favorite, 2nd favorite, and least favorite scales. **Results:** Scales B (8/30), D (8/30) and C (7/30) were preferred more often than scales A (3/30), E (3/30) and F (1/30) ($p < .1$). When 1st and 2nd favorite scale formats were combined, chi-square became highly significant ($p < .001$). Among the 3 favorite scales (B,C,D), no clients selected Scale D as their least favorite scale of the 6 formats evaluated. But Scales B (6/30) and C (11/30) were identified by some clients as their least favorite. Thus, based on clients' 1st, 2nd, and least favorite preferences, Scale D (qualifying words only, no quantification) was preferred by individuals with schizophrenia living in the community. **Conclusions:** Specifically, we have determined client preferences for item weighting scale formats for the SOAP-51. Generally, individuals have response format preferences that should be considered in the development of quality of life measures.

Abstract 1233**COMPARING THE QWB-SA, THE EQ-5D, THE SF-6D, AND THE NHP-D IN A MALE POPULATION SAMPLE**

Thomas Kohlmann, Sonja Boehmer, Franz Porzolt, Institute for Social Medicine, Medical University of Luebeck, Luebeck, Germany

Objective: To assess feasibility and psychometric properties of three utility-based instruments for measuring quality of life and the unidimensional version of the Nottingham Health Profile (NHP-D). **Methods:** A random sample of N=303 male residents of Luebeck, North Germany, completed a mailed questionnaire (age: 52-69, response rate: 64 percent). This comprised the recently translated self-administered version of the Quality of Well-Being Scale (QWB-SA), the EuroQol instrument (EQ-5D), the Short-Form Health Survey (SF-36), and the Nottingham Health Profile (NHP). Available scoring algorithms were used to produce index scores of utility-based scales and of the unidimensional version of the NHP. **Results:** Missing value rates on the item level were typically in the range of 1 to 3 percent for all quality of life measures. Score distributions were highly skewed in the direction of positive well-being. The EQ-5D and the NHP-D produced the most marked ceiling/ floor effects with 40 percent (EQ-5D) and 50 percent (NHP-D) of the respondents reaching minimum or maximum scale scores, respectively. Correlations of utility-based index scores with descriptive quality of life subscales of the SF-36 and NHP indicated that index scores were more closely associated with physical aspects of health (e.g., $r = .48$ to $.82$ with SF-36 physical component scale) than with psychological well-being (e.g., $r = .19$ to $.54$ with SF-36 mental component scale). Correlations between the utility-based instruments were moderate, ranging from $r = .35$ to $.65$ with a mean inter-scale correlation of $r = .50$. The NHP-D correlated with the utility-based measures in the same range. **Conclusions:** All instruments included in this survey appear to be feasible measures of quality of life in men between 50 and 70 years of age. However, the measures showed high skewness of score distributions. Moderate correlations between utility-based measures suggest that these cover only partially overlapping concepts of quality of life.

Abstract 1395**THE SENSITIVITY OF HRQOL MEASURES TO CHANGES AMONG ELDERLY INSTITUTIONAL RESIDENTS**

Walter P. Wodchis, Michigan Health Services Research Initiative, University of Michigan, Ann Arbor, MI

Most generic health-related quality of life (HRQoL) measures have been developed for community populations and the applicability of these measures to the institutional elderly population is not proven. The purpose of this study was to assess the validity and sensitivity of two existing utility-based HRQoL measures to evaluate the outcomes of elderly institutional residents who received psychotropic medications. The study population was all Vermont and Michigan nursing home residents with a 3 month stay between 1996 and 2000 ($n = 13,768$). Utility scores were assessed by mapping functional health status measured by the Minimum Data Set Resident Assessment Instrument (MDS) to domains of the Health Utilities Index Mark 2 (HUI2) and Mark 3 (HUI3). This study employed existing utility scores and did not directly elicit utility weights from the institutional elderly population. Resident changes in HRQoL scores over 3, 6 and 12 month intervals were stratified by medication use and clinical side-effects. Sensitivity was evaluated by significant changes in resident HRQoL scores and validity was assessed through agreement with hypotheses for changes in resident HRQoL scores. Mean (s.d.) HUI2 and HUI3 scores on admission were 0.25 (0.15) and -0.05 (0.20) respectively. Residents with side effects had lower HRQoL scores than those without. On follow up, average resident HRQoL showed decline in both measures. Controlling for treatment and side-effect profiles, both measures were sensitive to resident changes over time. Significant changes in HUI2 scores agreed with hypotheses while some HUI3 changes were counter-intuitive suggesting that the HUI2 is a more valid measure for this population. Very low HUI2 scores and negative scores on the HUI3 suggested that on average, residents were near or worse off than death indicating the need to use population-specific utility weights for the elderly institutional population.

Abstract 1694

ISSUES IN USING CONJOINT ANALYSIS TO VALUE HEALTH CARE WITHIN CLINICAL STUDIES

Isabelle Girod, Patrick Marquis, Mapi Values, LYON, FRANCE

Recent applications of conjoint analysis (CA) in health care showed the technique's potential for measuring benefits beyond traditional health outcomes, particularly patient preferences (PP). However, further methodological work is needed to generalize its application into clinical research. This abstract discusses the issues addressed during a recent PP research in benign prostatic hyperplasia (BPH) patients. Issues were addressed at the different stages of PP design: 1) identify attributes to include in the scenarios; 2) choose attribute levels; 3) select scenarios to present; 4) decide which was the suitable method to elicit patient preferences; and 5) choose the CA. 1-Attributes were identified using patient interviews, literature review and drug profiles (side-effect, daily cost, time to onset). The selection of the most important attributes for patients addressed the question of defining their relevance. Six attributes were selected in our study. 2-Attribute levels were determined to make trade-off possible while remaining realistic. The margin between these two concepts was narrow for the attributes chosen, for example the daily cost of medical treatment ranged from \$1 to \$1.5. 3-A fractional factorial design was employed to determine the 16 scenarios that were administered to keep the patient burden acceptable (< 12 minutes). Three sets of scenarios in different orders were administered, as order may affect the results of CA. 4-The method for PP elicitation was discussed in order to find a common design to assess quality of life, PP and clinical criteria within the same study. Auto-administration methods were preferred to interview-administration, to limit clinician burden and because of the sample size: 871 patients. Rating methods were chosen as the simplest approach. 5-Data were analyzed using ordered probit regression techniques. When following theoretical and methodological guidelines for assessing PP, researchers need to take into account study design, the size of the sample and the number of outcomes measured within the same study.

Abstract 1704

IMPACT3: RAPID DEVELOPMENT OF INTERNET-CAPABLE UTILITY ASSESSMENT INSTRUMENTS.

Leslie A. Lenert, Health Services Research and Development, Veterans Administration San Diego Healthcare System, San Diego, CA

The use of computers to elicit utilities may improve the precision and the validity of assessments, as well as reducing the logistic difficulty of performing assessments. However, a significant obstacle to the use of computer methods has been the cost and difficulty of creating software or web sites for utility measurement. The discipline of Medical Informatics provides important insights into how the process of design and implementation of could be automated. Utility elicitation protocols can be thought of as collections of named components that achieve specific tasks. For example, tasks in protocols often include collection measurement of health status or description of hypothetical states in addition to utility measurement. While the specific details of each component might change from study to study, the conceptual model is often similar. To design a protocol under this model, a researcher selects protocol components and then customizes each component to perform specific tasks required. This approach results in a complete and unambiguous specification of each step in the protocol that, in turn, allows computer program to automatically build software implementing it. We have developed a *free, publicly accessible* web site (<http://preferences.ucsd.edu/impact3.asp>) implementing these "informatics" principles. Using this site, researchers can design utility elicitation protocols using a web browser. Protocols can include instruments to collect demographic and health status data and can measure preferences for any number of states using any number of scaling methods (standard gamble, time trade-off, etc.). Once the design is complete, the web site generates a set of hypertext-markup-language files implementing researcher's protocol and downloads those files to the researcher's computer. The result is a (nearly completed) web site that a researcher can use in a "stand alone" fashion on a desktop computer or publish on a web server.

Abstract 1556

GETTING A GRIP ON MEASUREMENT ERROR: THE VALUE OF STRUCTURAL EQUATIONS WITH LATENT VARIABLES IN QOL RESEARCH

JA Scott-Lennox*, RD Lennox*, *Piedmont Research Institute, Chapel Hill, NC

Health-related quality of life (HRQL) studies often fail to detect reliable differences between treatments. Established statistical methods for controlling measurement error in self-report data can help overcome this limitation, but HRQL researchers have made limited use of these methods to date. Among the most versatile of these techniques is structural equation modeling with latent variables (SELV) which can enhance understanding of scaling, assessment validity, and test the relationship among disease processes, HRQL, and other treatment outcomes. This presentation will address these advantages conceptually so that non-statisticians can appreciate the value of SELV modeling. We introduce SELV with a discussion of confirmatory factor analysis (CFA) for evaluating multidimensional HRQL scales and for comparing translations. CFA will be contrasted to item response theory (IRT) models to highlight the strengths of each approach. Next, the presentation will show how SELV can be used to test hypotheses for treatment comparisons and evaluation of causal models of HRQL and other outcomes. Examples will highlight the importance of explicit measurement models for defining and testing relationships among HRQL measures and between HRQL and other disease or outcome indicators. Comparison to exploratory factor analysis, IRT, and more traditional regression analysis will demonstrate how SELV models control measurement error, and thereby reduce the risk of missing treatment effects without increasing the risk of false positive conclusions from scaling studies or treatment comparisons.

Abstract 1360

THE ASSOCIATION BETWEEN COGNITIVE FUNCTIONING AND THE RELIABILITY IN REPORTING PHYSICAL HEALTH RELATED QUALITY OF LIFE IN MULTIPLE SCLEROSIS PATIENTS

D. M. Miller, W. R. Lenderking, E. Tafesse, J. I. Greenstein, J. T. Phillips, Mellen Center, Cleveland Clinic Foundation, Cleveland, OH

Purpose: To determine the association between levels of perceived cognitive functioning (PCF) in MS patients and reliability and validity of self-reported physical functioning. MS produces mild to moderate cognitive deficits in about 50%, severe and global dysfunction in 5-10%. Monitoring health related quality of life (HRQoL) for patients affected by diseases characterized by cognitive impairment requires understanding how these deficits may affect the psychometric properties of HRQoL instruments. Subjects and methods: Generic and disease specific scores including PCF were collected from 2777 patients enrolled in the MSTRAC registry supported by Biogen. Physical wellbeing (PWB) was assessed using the SF-36 Physical Functioning (PF) & Role Physical (RP) scores. Patients in the highest and lowest quartiles of the Cognition sub-scale of the Modified Fatigue Impact Scale were characterized as high and low PCF. Internal consistency coefficients for the PF and RP were calculated separately in the two groups to assess reliability. Predictors of PF and RP were examined in the two groups in regression models examining disease severity, psychological well being and pain. Results: Patients with lower PCF reported both PF and RP with significantly less reliability, although only the RP coefficient was less than expected for the general population. Strikingly, considerably less variance in both PF and RP was predicted in the worse PCF group. Conclusion: As expected, the psychometric properties were affected by PCF with patients with lower levels of cognitive functioning showing less reliability and predicting less of the variance in aspects of physical functioning. This preliminary evidence suggests that cognitive impairment levels may need to be controlled for in examination of HRQoL in some patient groups. Future research using objective measures of cognitive functioning should be performed to confirm these findings.

Abstract 1666**COMPARISON OF PATIENTS AND INVESTIGATORS GLOBAL RATING OF SEVERITY AND OF CHANGE**

Patrick Marquis, Alain Follet, Dominique Dubois, Margareth Rothman, Mapi Values, Lyon, France

Likert scales are often used to assess global severity of disorders and global changes after treatment. It is commonly said that investigator and patient ratings may differ. But exact agreement, over-estimation or under-estimation by investigators, as opposed to patients, is not always quantified. The objective was to compare patient and investigator global ratings of severity and change. A 7-point Likert scale (from absent to very severe) was used to assess global severity of GI disorders in a 2-week observation plus a 4-week usual care treatment study. The scale was filled in at inclusion, week 2 and week 6 by 223 US patients and their investigators. At week 6, an other 7-point Likert was filled in to assess change overtime (from very much improved to very much worsened, level 4 being unchanged). Agreement has been analyzed using bubble plots and weighted Kappa coefficient. A perfect agreement was found for the severity Likert scale in only 38%, 34%, and 44% at week 0, 2, and 6 respectively. Investigators underestimated the severity compared to patients in 33%, 28%, and 21% at the 3 visits (over estimation by investigators was 29%, 38%, and 35%). Regarding the Likert scale assessing change overtime, the agreement was slightly better: 56% at week 6. Investigators underestimated the change compared to patients in 21% (over estimation of change by investigators was 23%). The weighted kappa of 0.34, 0.28, and 0.50 at the 3 visits represented poor agreement between investigators and patients for the evaluation of severity. The weighted kappa was 0.72 at week 6 represented good agreement beyond chance for the assessment of change. Bubble plots for severity at week 0 and week 6, and assessment of change at week 6 illustrate the results. The disagreement is higher regarding assessment of severity than assessment of change between investigators and patients. For perceptual evaluations, patients' perspective should be preferred.

Abstract 1051**USING PAPER-AND-PENCIL SURVEYS ON THE WEB TO ASSESS QOL; SHOULD WE EXPECT DIFFERENCES IN RELIABILITY?**

David Litaker, Penny Ott, Linda Muscatello, The Cleveland Clinic Foundation, Cleveland, OH

Introduction: Web-based surveys are commonly used to access large populations at minimal costs in obtaining research data. Although many factors make this method of data acquisition attractive, the reliability of responses, especially in assessments of latent constructs such as quality of life (QoL), is unclear even if previously validated paper-based surveys are used without alteration. **Methods:** The Rhinoconjunctivitis Quality of Life Questionnaire (RQLQ) and the allergy-specific Work Productivity and Activity Impairment (WPAI) were formatted into a single paper-and-pencil form (P) and two computer-assisted interfaces: touch screen (TS), and web-based (WB). Employees at a tertiary care medical institution completed the survey sequentially in two of the three formats, assigned randomly and in random order. Responses for each individual were compared on an item-by-item basis using Pearson's correlation and kappa coefficients; composite scores were compared within groups using paired T tests and between groups using one way ANOVA. **Results:** 75 volunteers completed 150 questionnaires in one of three possible format pairs (P-TS, P-WB, and TS-WB). 54 (72%) of respondents were female; median age was 40-49 years. Individual item responses were highly correlated and concordant for all formats ($p < 0.001$). Between group comparisons of the composite score calculated from responses in the RQLQ or WPAI showed no significant differences in QoL or work impairment due to allergies among administration formats. Within group comparisons, however, demonstrated significant differences between paper- and internet-based forms in allergy-related work impairment ($p < 0.001$, corrected for multiple comparisons). **Conclusions:** Although response concordance and correlation is similar, important differences in reliability may arise when different formats of administration are used. This study re-emphasizes the importance of establishing instrument reliability when new survey technology is used.

Abstract 1568**CONFLICTING RESULTS AND LEVEL OF ANALYSIS: A LONGITUDINAL STUDY OF ALOPECIA**

James P. Donnelly, Susan Nolte, Sharon Kelly, Patricia Conley, Roberta Cobb, GOG Statistical Office, Roswell Park Cancer Institute, Buffalo, NY

Studies of the psychological impact of alopecia resulting from chemotherapy are conflicting. One explanation for differing outcomes is the level of analysis. The present study used a longitudinal design and two levels of analysis to examine adaptation to hair loss in women treated for pelvic malignancies in one of 14 clinical trials of the Gynecologic Oncology Group. One hundred and ninety women who met eligibility criteria were given the Body Cathexis Scale at three time points. The scale contains 40 body-image and 40 self-image items, resulting in two global scores. Degree of hair loss was rated by nurses on a three level scale (none, mild, pronounced/total) at each time point. By the second assessment, 88% of the women had experienced the most severe alopecia and by the third assessment nearly all had reached this level. In order to examine whether level of analysis determines the pattern of results, two approaches to scoring the data were used. In the first analysis, means were computed separately for body and self image and these scores were tested in a repeated measures analysis of variance. There were no significant differences over time for either of the two global scales. The second set of analyses were based on factor analysis of the baseline body and self image data. The analysis of the body image scores included a five item facial appearance factor which was deemed theoretically most sensitive to the experience of alopecia. Facial appearance factor scores showed a significant decline from baseline to the third assessment ($F = 3.897, p = .021$). A self-confidence factor, which was thought to possibly mediate the negative impact of alopecia, showed a significant positive change over time ($F = 3.639, p = .027$). These results suggest that level of analysis should be based on specific theoretical constructs to avoid misinterpretation, and that adaptation to side effects may often involve both positive and negative changes over time.

Abstract 1216**INTERPRETATION OF QOL-CHANGES IN INPATIENT REHABILITATION PROGRAMS FOR CANCER PATIENTS**

Michael T. Moser, Joachim Weis, Dorothee Fachinger, Hans Helge Bartsch, Institute for Rehabilitation Research, Tumorbiology Center, Freiburg, Germany

Clinical significance is a major issue in the discussion about the meaningfulness of QoL-changes. Several concepts were outlined in the literature. Medical and psychosocial rehabilitation programs for women with breast cancer were recognized as important interventions because positive effects on QoL, relapse and survival were reported. 161 women with non-metastatic breast cancer were included. QoL was assessed by the EORTC-QLQ-C30/breast module at admission and discharge (approx. four weeks later). All women participated in interdisciplinary medical and psychosocial treatment programs. Changes in QoL were analysed with MANOVA procedures. Four concepts of clinical significance were compared: (1) effect size statistics, (2) reliable change statistics, (3) subjective significance measures and (4) normative comparisons. Statistical significant improvements in all functional domains and significant decreases in most symptom scales were found ($p < .02$). Effect size calculations show large effects in the domains emotional functioning (EF) and general QoL (GQoL) and moderate effects in the domains physical (PF)-, role- and social functioning (SF), fatigue, breast symptoms and therapy side effects (according to Cohen's classification). Reliable change statistics show reliable GQoL-improvements in 52% contrasted with 6% reliable deteriorations. A crossvalidation of QoL-changes with patient's subjective significance self-ratings indicate large improvement in EF (mean change score/MCS= 20.3 points) and moderate improvements in PF (MCS= 3.8), SF (MCS=13.3) and GQoL (MCS=13.2). Normative comparisons and equivalency tests provide information about the comparability of QoL-profiles between our sample and women in general population. Methods for determination of clinical significance are an important dilation of classical hypothesis-testing procedures in QoL-research. Our findings suggest strong evidence for the efficacy of oncological inpatient rehabilitation programs.

Abstract 1244

MINIMAL CLINICALLY IMPORTANT DIFFERENCE THRESHOLDS AND THE STANDARD ERROR OF MEASUREMENT: IS THERE A CONNECTION?

Kathleen W. Wyrwich, School of Public Health, Saint Louis University, St. Louis, MO

Several recent investigations examined the relationship between the standard error of measurement (SEM) and established criteria for minimal clinically important difference (MCID) scores on latent constructs of health status. These investigations, however, have resulted in differing SEM criteria for the MCID. This study reviews the investigations linking a SEM-based criterion to an established MCID threshold. Two investigations using disease-specific health-related quality of life (HRQoL) measures among patients with the chronic diseases (coronary artery disease, congestive heart failure, or chronic obstructive pulmonary disease) have consistently found that one SEM corresponds to the established MCIDs for HRQoL domains. Similarly, three investigations among patients with back and neck pain who were referred to physical therapists found that 2.3 SEMs consistently approximates the established MCID standards for measures of pain. In addition to the SEM-based criteria convergence for the MCID within these conditions, these investigations revealed a connection between the 1 and 2.3 SEM criteria. A minimal clinically important difference occurred among the chronic disease patients if the patient rated their change (improvement or decline) at 1, 2, or 3 (almost the same, hardly any better or worse; a little better or worse; or somewhat better or worse) on a 7-point Likert scale, although few patients chose a rating of 1. The back and neck pain patients, however, needed a change rating of 5, 6, or 7 (a good deal better, a great deal better, a very great deal better) on the same Likert scale to experience an MCID. Charting these change levels against their respective SEM-MCID criteria provides promise for linking a SEM-based criterion to an MCID standard in other latent constructs of health status.

Abstract 1401

CLINICAL MEANINGFULNESS (CM) OF DIFFERENCES IN HEALTH-RELATED QUALITY-OF-LIFE SCORES IN CLINICAL TRIALS.

David Osoba, Department of QOL Consulting, University of British Columbia (Professor, Ret.), West Vancouver, BC, Canada

Tests of statistical significance are dependent on sample size. Small numerical differences in health-related quality-of-life (HRQL) scores may be statistically significant when the sample size is large, but not if the sample size is small. Also, the magnitudes of numerical differences that are CM may vary with various users of HRQL information, i.e., patients, clinicians, and policy setters. Thus, in addition to statistical significance, tests of CM are required to interpret changes and differences in HRQL scores. Recent studies show examples of analytic approaches that address these issues. A comparison has been made between statistically significant results and the results of other approaches, i.e., effect size and determination of CM to patients (subjective significance) enrolled in oncology clinical trials. The trials consisted of various chemotherapy regimens in breast cancer, small-cell lung cancer, hormone-resistant prostate cancer and malignant gliomas (Osoba D et al, J Clin Oncol 1998;16:139, 1999;17:1654, 2000;18:1481). When patients with breast or lung cancer rated the degree of change that they perceived in certain domain scores (EORTC QLQ-C30) over time, a small change was associated with a change between 5-10 points, a moderate change with 11-20 points and a large change with >20 points (scale 0-100). Effect sizes were about 0.2-0.5, 0.5-0.8 and >0.8, respectively. Studies in malignant gliomas have shown small to moderate effect sizes despite P values not being statistically significant (>0.05). Using a 10-point change in HRQL scores lasting for at least two treatment cycles as an endpoint, patients can be classified into those with improvement, no change and deterioration in HRQL scores. The proportions of patients achieving these endpoints has been shown to be significantly different between arms in clinical trials. It is concluded that methods for calculating CM provide useful information in addition to that provided by tests of statistical significance.

Abstract 1404

MEASURING PATIENT PERCEPTION OF HEALTH-RELATED QUALITY-OF-LIFE CHANGES USING THE SUBJECTIVE SIGNIFICANCE QUESTIONNAIRE.

George Rodrigues, Andrea Bezjak, David Osoba, Pamela Catton, Debbie Tsuji, Diane Taylor, Pdraig Warde, Department of Radiation Oncology, Princess Margaret Hospital, Toronto, ON, Canada

Objective: To examine the relationship between changes in health-related quality-of-life (QOL), as measured by the EORTC Quality of Life Questionnaire (QLQ-C30), and the patients perception of change in their QOL during radical radiotherapy (RT) for cancer of the prostate. Methods: A total of 101 patients completed the QLQ-C30 on weeks 1, 4 and 7 of radical external-beam RT for localized (T1-T3NOMO) cancer of the prostate. Patients rated their change in physical, emotional, social, and overall QOL by completing a seven-category subjective significance questionnaire (SSQ) at weeks 4 and 7. The association between the QLQ-C30 change scores and the corresponding SSQ ratings were determined by Spearman rank correlation coefficient analysis. Results: Patient compliance with the QLQ-C30 and SSQ exceeded 95 percent. In the various QOL domains, between one-third and three-quarters of patients had perfect baseline QLQ-C30 QOL scores (QOL score = 100). Statistically significant changes in fatigue, pain, appetite, diarrhea, and overall QOL QLQ-C30 scores were detected during RT. For patients reporting "a little" change in QOL (the "subjectively significant" change), mean QLQ-C30 change scores generally ranged between 2.5 to 8.5 points. In the entire study sample, correlations between SSQ patient ratings and QLQ-C30 change scores were lower than previously reported, ranging between 0.15 and 0.24 for the four different QOL domains. However, correlation improved (0.12-0.46) when patients with perfect baseline QLQ-C30 scores were excluded from the analyses. In addition, patient perception of QOL change correlated best with changes in the QLQ-C30 pain score. Conclusions: The SSQ is a useful tool for the proper interpretation of clinically relevant changes in QOL, although the strength of correlation is dependent on baseline patient characteristics.

Abstract 1416

INTERPRETING THE SIGNIFICANCE OF CHANGES IN EORTC QLQ-C30 AND COOP/WONCA SCORES

Kommer Sneeuw, Martin Muller, Neil Aaronson, Division of Psychosocial Research and Epidemiology, The Netherlands Cancer Institute, Amsterdam, The Netherlands

The purpose of this study was to define relevant changes in mean QoL scores assessed by the EORTC QLQ-C30 and the COOP/WONCA charts using the minimal clinically important difference (MCID) method. We also examined the viability of estimating proportions of patients who deteriorate, remain stable or improve over time. A heterogeneous sample of 240 cancer patients receiving chemotherapy completed the two instruments at an early phase of treatment (T1) and three months later (T2). For six joint domains, transition questions were presented to the patients at T2 assessing their retrospective perception of change. Correlations between prospective and retrospective change scores were rather weak (average $r=0.39$, range 0.24-0.47). Closer examination of the change scores showed that, when using retrospective ratings, patients failed to discriminate between the different functional domains. That is, high correlations between these domains were found (average $r=0.70$, range 0.64-0.83). Also, the retrospective change scores were associated with T2 scores only. Classification of patients into those feeling worse, stable or better proved difficult when using EORTC QLQ-C30 scores, because the scales are based on different numbers of items and response options. The clear structure of the COOP/WONCA charts allowed a consistent classification. We conclude that transition questions, though frequently used and seemingly straightforward, may not be valid. Patients appear to have difficulties in discriminating between different functional domains and in remembering their previous condition. These findings suggest that retrospective questions may not be appropriate for defining relevant changes in QoL scores. Classification of patients into those who deteriorate, remain stable or improve is difficult for multi-item scales. An instrument using one single item for each QoL domain, and employing a limited number of clearly labeled response options, may be more appropriate for this purpose.

Abstract 1565**QUALITY OF LIFE OF GERIATRIC PATIENTS AFTER PROXIMAL FEMORAL FRACTURES TREATED BY JOINT REPLACEMENT OR OSTEOSYNTHESIS**

Hans Josef Erli, Victoria Fernandez, M. Marx, Mark Brüggmann, Ottmar Paar, Department of Trauma Surgery, University Hospital of the RWTH Aachen, Aachen, Germany

The aim of this study was to determine physical outcome and Quality of Life after fractures of the proximal femur in old-age patients and different forms of treatment. We examined 51 female patients with a mean age of 73.7 years, ranging from 60 and 93 years. They all had suffered from a fracture of the proximal femur. 49 of all patients were treated surgically, in 2 cases conservative treatment was possible. Of the operated patients 49 % were treated with a special implant like DHS or Gamma-Nail, joint replacement was performed in 40,8 % (hemiarthroplasty in 20,4% and total replacement in 20,4%). Other implants were used in 10,2 %. The physical outcome was reported according to the hip-scores of Merle d'Aubigne', Harris, Iselin and Charney. For the assessment of Quality of Life we used a Visual Analogue Scale (VAS), the Spitzer Index, the NHP, the ELQ, and the reALOS. The assessment of physical outcome after proximal femoral fracture shows in all hip-scores that patients treated by joint replacement reached significantly better physical outcomes and lower pain levels (from $p=.002$ to $p=.047$). The Visual Analogue Scale (VAS) as a valid measurement for global Quality of Life shows a significant difference between patients treated by joint replacement and those treated conservatively or by reconstructive surgery ($p=.028$). The other QoL-questionnaires support the results of the VAS in a significant way. This study shows, that physical outcome and QoL after proximal femoral fractures of old-age patients is related to the surgical procedures performed. According to our results these patients profit from joint replacement, whereas those who had undergone reconstructive surgery had more pain and reached lower QoL ratings. Although if other variables such as the fracture type must be considered, these results are relevant for the planning of surgical procedures.

Abstract 1085**USING THE SF-36 WITH OLDER ADULTS: CROSS-SECTIONAL COMMUNITY BASED SURVEY**

Stephen J. Walters, James F. Munro, John E. Brazier, School of Health & Related Research (ScHARR), University of Sheffield, Sheffield, United Kingdom

Although people over 65 are a growing proportion of the population and major users of health care, they have frequently been neglected in both the development and use of health-related quality of life measures. This may be because of uncertainty over the practicality and validity of using such measures with older adults. The aims of this study were to establish whether the SF-36, a popular self-completed health status instrument, could be successfully used in a postal survey of older adults, and to provide population scores for a representative sample of community-dwelling adults over 65 years old. We undertook a survey of all 9897 adults aged 65 to 104 registered with twelve general practices in Sheffield, UK. Each was sent a questionnaire booklet containing the SF-36 and other health related items. Non-respondents received up to two reminders at three-weekly intervals. The main outcome measures were scores on the eight dimensions of the SF-36 and a modified version of the physical functioning dimension. The SF-36 achieved a response rate of 82% ($n=8117$) and dimension completion rates of 86.4% to 97.7%. Internal consistency measured by Cronbach's alpha exceeded 0.80 for all dimensions except social functioning. These results compare favourably with postal surveys of younger adults. Scores for older adults were calculated by age and sex. Comparison with data from a similar survey of younger people in the same city showed how functional dimension scores decline steeply with age, in marked contrast with mental health. We conclude the SF-36 is a practical and valid instrument to use in postal surveys of older people living in the community.

Abstract 1107**WHAT DO OLDER PEOPLE EXPECT FROM HEALTH AND SOCIAL CARE IN THE COMMUNITY?**

Petra Kliempt, Danny A. Ruta, Simon A. Ogston, Marion E. McMurdo, Department of Epidemiology and Public Health, University of Dundee, Dundee, Scotland, Great Britain

A survey was conducted to identify the outcomes of community care for people aged 75 years and over considered most important for their quality of life; and to compare responses between recipients of care, and health and social work professionals. Initial pilot interviews were conducted with 10 older people aged between 75-97 years, living in their own homes, to identify the most common desired outcomes. An unstructured interview approach was used and people were asked about the health and social care they had been receiving and about factors they considered most important and relevant to their quality of life. These outcomes were then incorporated into a postal questionnaire in which we invited older people ($n=165$) as well as the health and social care professionals ($n=157$) to rank them in order of importance. Pilot interviews generated six desired outcomes: company, mobility, pain relief, personal hygiene, safety and social support. Response rates of 32% from older people (mean age=80 years) and 64% from health and social work professionals were achieved. The option was given to return the questionnaire blank if they did not wish to take part. 40% of older people returned the questionnaire blank, and 33% of those gave reasons for not completing it. Common reasons given were: being in good health, family and friends provide care, and that the questionnaire was not relevant to their needs. The survey showed that there were few differences between older people and health and social care staff regarding the ranking of the six outcomes. Company and mobility were the most highly ranked outcomes and personal hygiene the least. Neither staff nor patients suggested any outcomes other than the ones listed. When evaluating health and social care in the community any package of measures must at least provide an assessment of the six outcomes identified in this study.

Abstract 1094**HEALTH AND OTHER ASPECTS OF THE QUALITY OF LIFE OF OLDER PEOPLE**

Alex C. Michalos, Anita Hubley, Bruno D. Zumbo, Dawn Hemingway, Institute for Social Research and Evaluation, University of Northern British Columbia, Prince George, BC, Canada

In September 1999 a survey was undertaken throughout the Northern Interior Health Region of British Columbia. A total of 875 people completed 23-page questionnaires, the average age was 69 and the range 55 to 95. Responses to the SF-36 indicated that for male respondents 55-64 the mean score for the 8 dimensions was 74.4, practically identical to the US norm for this age group and lower than that of the UK. For males 65+ the mean was 68.3, practically the same for the US norm. For females 55-64 the mean score was 73, better than the US norm of 70.6. For females 65+ the mean score was 65.4, equal the US norm. Comparing 18 average figures for these older people on satisfaction with specific domains of life and life as a whole with average adults in P.G. in Nov. 1999, in all but 2 cases the older people's scores were higher. Eleven percent of our respondents had been a victim of a crime in the last year, compared to 38% in a 1997 P.G. survey of all adults. Two or three of the 8 SF-36 health dimensions explained 37% of the variation in life satisfaction scores, 34% in happiness, 34% in satisfaction with the overall quality of life and 22% in satisfaction with one's overall standard of living. Using SF-36 scores combined with worries about crime and other aspects of people's lives, we could explain 60% of the variance in life satisfaction scores, 44% in happiness, 58% in satisfaction with the overall quality of life and 59% in satisfaction with one's overall standard of living scores.

Abstract 1575

MEASURING QUALITY OF LIFE IN MODERATE TO SEVERE COGNITIVE IMPAIRMENT: ARE PROXIES ALWAYS NECESSARY?

Marlene A. Reimer, Carole-Lynne LeNavenec, Susan Slaughter, Faculty of Nursing, University of Calgary, Calgary, AB, Canada

Proxies are commonly used when patients have cognitive impairment. There is little evidence to guide researchers as to how these patients can contribute to evaluating their own quality of life (QOL). We report on 3 studies in which we have pushed the limits on self-report. STUDY 1: The psychometric properties of the Comprehensive QOL Scale (Intellectual Disability Version), the Sickness Impact Profile, and the SF-36 were tested in 152 adults with cognitive impairment from traumatic brain injury (TBI), schizophrenia, multiple sclerosis, or obstructive sleep apnea, and their proxies in an interviewer administered format. Proxy scores correlated at .44, .61, and .63 respectively. Subjects with <5 scores in the severe range on the Neurobehavioral Cognitive Status Examination could respond to these measures but not those with 5 or more severe scores. STUDY 2: Using the critical incident technique (CIT) we found survivors of TBI with Rancho scores of 6-7 (and some 5s) were able to describe incidents which showed what was important to their QOL. By using CIT with family members, volunteers, and staff we obtained additional incidents, many of which included non-verbal behaviors of survivors seen as indicators of QOL. STUDY 3: In a study of QOL and cost of care for residents with middle to late stage dementia we used a composite of objective and subjective measures including the Pleasant Events Scale, Cohen-Mansfield Agitation Inventory and Apparent Affect Scale to essentially "read" the non-verbal evidence these residents provided about their state of well-being. In summary we have shown that patients with moderate cognitive impairment can participate in structured interviews and those with severe impairment can communicate their experience of QOL through non-verbal behavior. Thus the need to rely on proxies can be reduced and when input from significant others is required it can be based on observed behaviors rather than opinion as to how the affected person might respond.

Abstract 1514

PROXY USE OF THE FRENCH SF-36 AND PDQ-39 IN RATING QUALITY OF LIFE OF PARKINSONIAN PATIENTS

Nathalie Andreu, Jean-Paul Collet, James Hanley, Jean-Louis Montastruc, Olivier Rascol, Sharon Wood-Dauphinee, Departments of Epidemiology and Biostatistics, McGill University, Montreal, Canada

The use of proxy raters of patients' quality of life (QOL) has been advocated as a means of avoiding missing data in QOL studies. This source of information could become crucial in Parkinson's disease (PD) since PD patients may experience growing physical and cognitive problems, which limit their ability to self-rate questionnaires. In order to study the accuracy of proxy information on patient's QOL, we examined the degree of agreement between the PD patient and his/her caregiver when completing the SF-36 and the PDQ-39, a PD-specific QOL questionnaire. This investigation was conducted in a series of 126 non-demented PD patients from our prospective database. We asked patients to ensure that a family member or relative who knew them well was available at the time of the interview. One neurologist and two psychologists rated the motor and mood status of the 126 PD patients, attending the neurological department, with traditional clinical scales. Then, PD patients and 110 proxies self-completed the SF-36 and PDQ-39 on 2 separate occasions, during the same day. For SF-36 and PDQ-39, agreement in patient-proxy pairs ratings was best for objective physical dimensions ($ICC > 0.60$), moderate for psychological dimensions ($0.40 < ICC < 0.61$) and worst for social dimensions ($ICC < 0.41$). Agreement was better for PDQ-39 than for the corresponding SF-36 dimensions. Patient-proxy agreement was lowest on all SF-36 dimensions for more severe PD patients. At the group level, proxies systematically underestimated patients' QOL on most PDQ-39 dimensions but this difference remained small ($d < 0.30$). Further, the mean scores and standard deviations of patient and proxy samples were fairly similar. Although, at the individual level, proxies are far from optimal raters of patients' QOL, proxy ratings may be used for making group comparisons. However, the use of proxy data intermingled with data from PD patients to measure QOL may lead to biased results.

Abstract 1799

APPLICATION OF THE EXTENDED Q-TWIST METHODOLOGY AMONG MIGRAINEURS: RESULTS FROM A CLINICAL TRIAL

Shoshana S. Colman, Betsy A. Rothermich, Clayton R. Rowland, Baltazar Gomez-Mancilla, The Lewin Group, San Francisco, CA

The objectives of this research were to measure individual preferences among patients experiencing a migraine headache and to incorporate the values into an EQ-TWiST (Extended Quality-adjusted Time Without Symptoms and Toxicities) analysis. An EQ-TWiST analysis incorporates both the duration and quality of health states for patients. The EQ-TWiST has not been applied in the area of migraine and the measurement of patient preferences in clinical trial settings has been limited. 1255 otherwise healthy adult patients with migraines (based on the International Headache Society criteria) were enrolled in a double-blind, randomized, two-arm, parallel-group study comparing oral almotriptan 12.5 mg vs. oral sumatriptan 50 mg for the course of one migraine attack. Prior to the onset of a migraine, patient preferences were assessed for various health state attributes, including functional status, migraine pain, migraine symptoms, relevant quality of life domains, and work loss. The patient preferences and attribute scores were used to create a weighted assessment score (WAS) at 7 time points over the 48 hour study period. The WAS ranged from 0 (lowest quality of life) to 1 (highest quality of life). Then, an EQ-TWiST score was calculated as the area under the curve of the WAS (plotted over time). The EQ-TWiST score may be interpreted as a measure of quality adjusted life hours. At 24 hours following migraine onset, the mean EQ-TWiST score was 18.10 for the almotriptan group and 18.28 for the sumatriptan group (with 24.00 representing the ideal). After 48 hours, the mean EQ-TWiST score was 38.04 for the almotriptan group and 39.32 for sumatriptan group (with 48.00 representing the ideal). These results indicate that the EQ-TWiST methodology is sensitive to the impact of migraine. In addition, this research demonstrates the successful measurement of preferences and application of the EQ-TWiST methodology in a clinical trial setting for an acute episode of a chronic condition.

Abstract 1182

COGNITIVE INTERVIEWING: A METHOD FOR UNDERSTANDING THE MEDIATORS OF RESPONSE SHIFT?

Joanne Greenhalgh, Amanda Georgiou, Andrew F. Long, Keith Hurst, Nuffield Institute for Health, University of Leeds, Leeds, West Yorkshire, UK

This pilot study used cognitive interviewing to explore how people with multiple sclerosis calibrate the severity and impact of their symptoms in response to a quality of life diary in order to further understand the process of response shift. Nine people with MS were asked to think aloud and explain their answers as they completed one day of an adapted version of the quality of life diary used by Parkin et al (2000). Respondents calibrated their judgements of the severity and impact of their symptoms through a process of social comparison with several, often conflicting "benchmarks". These included "self with MS" either during a relapse or feeling well; "self previous to MS", an "ideal self" that they would like to be, the general population and other people with MS who were "worse off" than themselves. Respondents reported that they had shifted their expectations of their health and what they wanted to do as a result of having MS. This was not a one off process. Respondent's benchmarks were dynamic and had changed in response to their disease trajectory. These shifts in benchmarks for comparison had a significant influence on the way the response scales were calibrated and consequently the answers given by respondents to the questions in the diary. Respondents reported that the use of the different benchmarks to calibrate the severity of current symptoms and their impact would give a different answers to the same questions. Although the study did not explore response shift directly, it uncovered a wealth of information about the mediating factors that might underlie this process. It is suggested that one method through which response shift occurs is a change in the benchmarks used in the social comparison processes employed in calibrating response scales. Cognitive interviewing techniques are valuable methods of accessing and exploring the mediators and process of response shift.

Abstract 1596

PATIENT-SPECIFIC MEASURES OF QUALITY OF LIFE IN A READILY ACCESSIBLE DATABASE: PART OF MS.COMP, A COMPREHENSIVE CARE PROGRAM FOR MULTIPLE SCLEROSIS
 William H. Likosky, Howard Barkan, Jeff Klingman, Jay H. Rosenberg, Allan L. Bernstein, Jack Burks

MS is a chronic neurological disease with diverse symptoms and variable disease progression. Despite there being disease modulating agents, emphasis remains on aiding individuals' adaptation to the effects of the disease on medical and social-psychological domains. The most effective current interventions address symptoms resulting from target organ damage and the symptoms' effects on function and quality-of-life. Variation among providers in the assessment and provision of care may result in suboptimal support of MS patients. We are now implementing MS.COMP, a multisite program of distributed centers of excellence for MS care at 5 Kaiser-Permanente facilities and a multi-specialty group practice to reduce inter-provider variability and increase adherence with evidence-based guidelines. The goal of the program is to provide a consistent process of care, including all MS patients treated at these facilities regardless of their attending physician. Key components of MS.COMP include a database built around a series of encounter forms, allowing the recording of patient-specific data from multiple clinicians in various settings using identical formats, and around patient and clinician completed measures of symptoms, functional status, and quality of life (QoL). The database aids the systematic collection of clinical information over time via paper forms or direct entry. It gives clinicians ready access to clinical data. The QoL measures allow clinicians to judge the effectiveness of their symptom-specific and general interventions. This technique is applicable to other chronic diseases. We will present the database, the QoL measures, the mechanisms for entering and reporting them from the database, and analyses of the pilot data regarding the QoL measures and their associations with symptoms.

Abstract 1748

THE MEDICAL OUTCOMES STUDY 36-ITEM SHORT-FORM HEALTH SURVEY (SF-36): EXAMINATION OF SCALING ASSUMPTIONS IN PEOPLE WITH MULTIPLE SCLEROSIS (MS)

Jeremy C. Hobart, Jenny A. Freeman, Donna L. Lamping, Ray Fitzpatrick, Alan J. Thompson, Neurological Outcome Measures Unit, Institute of Neurology, London, England, UK

The SF-36 is a widely used generic health outcome measure. Two types of scores can be generated: Likert's method of summated ratings is used to produce eight multi-item scale scores; these are weighted and combined using a specific scoring algorithm to generate physical (PCS) and mental (MCS) summary scores. The latter are preferable for clinical trials. Although, the scaling assumptions underpinning these scores have been shown to be valid in a variety of clinical populations they have not yet been tested in MS. SF-36 data from 438 people representing the full spectrum of MS were examined (70% female, mean age 48 years). There was clear support for the generation of scale scores (e.g. symmetrical item response distributions, equivalent variances and item-total correlations, > 96% definite scaling success rates). However, only partial support was demonstrated for the generation of summary scores. Although a 2-component model of health was supported, the components explained < 60% of the total variance in SF-36 scales and < 75% of the total variance in four scales. Also, component-scale correlations did not fully support the use of a scoring algorithm based on US general population data. Intercorrelations between scales suggested this was because relationships between the eight health concepts measured by the SF-36 are disease-specific. Results confirm that when used as a health measure in MS, SF-36 scale scores can be reported with confidence but SF-36 summary scores should be used with caution. Results also suggest that it may not be possible to generate MS specific algorithms.

Abstract 1093

THE EFFECT OF RESPONDENT CHARACTERISTICS ON STANDARD GAMBLE VALUATIONS OF SF-6D HEALTH STATES

John Brazier, Jennifer Roberts, Sheffield Health Economics Group, University of Sheffield, Sheffield, UK

An important concern in quality of life measurement is whether preferences for health of individuals differ systematically across groups defined by characteristics such as age, sex and educational status. This paper investigates this issue using the results of a survey of the UK general population designed to elicit valuations of different states of health. Health states are defined according to the SF-6D; a new instrument derived from the SF-36 in order to facilitate estimation of a single index measure of health benefit for use in economic evaluation and population health surveys. The SF-6D has six dimensions each with between four and six levels; it defines 18,000 possible health states. A sample of 250 states have been valued by a representative sample of 579 individuals using an interviewer administered version of standard gamble. The overall aim is to construct a model for predicting health state valuations for all the states defined by the SF-6D. The econometric modelling must cope with the hierarchical nature of the data (each respondent values 6 health states) and its skewed distribution. The models have produced robust estimates of the 'main effects' and, in general, the results support the ordinality of the SF-6D scales. The results suggest that while the characteristics of each health state are the main factors determining the valuation given to that state, standard gamble valuations also vary according to the age and educational status of the respondent. There is also some indication that the extent to which the respondent understood the standard gamble task has a systematic effect on the health state valuation. Valuations tend to increase with age to around 60 years and then decrease in a quadratic pattern. Respondents who were educated to degree level and who have a good understanding of the standard gamble task also tend to give higher valuations.

Abstract 1201

ARE PREFERENCE SCORES ELICITED WITH THE VISUAL ANALOGUE SCALE INAPPROPRIATE FOR CALCULATING QALYS?

Paul F. Krabbe, Paul Kind, Frank T. de Charro, Department of Medical Technology Assessment, University Medical Centre Nijmegen, Nijmegen, The Netherlands

There is still controversy about the relationship between preference scores evoked with Visual Analogue Scales (VAS) and preference scores elicited with methods based on tradeoff concepts. Health economists, referring to axiomatic theory, postulate that only preference scores elicited with the standard gamble and the time trade-off (TTO) method are valid for computing quality-adjusted life years (QALYs). After a brief introduction into this ideological field, results from a large empirical study will be presented that comprised a head-to-head comparison of these two different methods. A representative sample of the general population (n = 2,997) valued 42 health states using both VAS and TTO. Respondents were interviewed in their own homes by trained interviewers. Health-state descriptions were based on the EuroQol-5D classification. The specific relationship between preference scores obtained with the VAS and the TTO method was studied in detail by performing several transformations of the VAS data. In addition, the topic how to deal with health states valued by individuals as worse than death was studied. The mean scores for the 42 health states produced by the two methods appeared to be highly similar once a single parametric power transformation has been applied to the VAS ($TTO = 1 - [1 - VAS]^a$). The parameter of the power function (a) was 3.09. The percentage of variance explained by the concave power function was high (0.97). Setting all individual scores for health states valued worse than death to zero hardly had any effect on the results. It seems that the simple valuation task embedded in the VAS evokes preference scores that, after a simple power transformation, are similar to those obtained using TTO. This implies that the highly feasible VAS can be used as a substitute for the TTO, substantially simplifying one important step to arrive at QALYs.

Abstract 1309**ASSESSMENT OF THE MULTI-ATTRIBUTE PREFERENCE FUNCTION FOR HUI 3 IN FRANCE**

Catherine Le Galès, Catherine Buron, Nathalie Costet, Sophia Rosman, Gérard Slama, CHEAR U537, INSERM, Le Kremlin-Bicêtre Cedex, France

Objective : The Health Utilities Index is a generic multiattribute preference-based system for assessing health-related quality of life, originally devised by Torrance et al. It is being widely used in economic evaluations in North America and in international multicentre studies but was not available in France. Following cross-cultural adaptation and psychometric validation of the HUI3 classification in the French population, the purpose of this investigation was to derive French population weights. **Methods :** As the HUI3 consists of 8 attributes deliberately selected to be structurally independent and specifies 5 or 6 levels by attribute, the explicit approach of the multi-attribute utility theory provides an efficient framework to determine the utility scores of the 972,000 health states of this system. In June 1999, we implemented a population (n = 365) preference survey. Respondants were randomly selected by phone to be representative of the French general population and face-to-face interviews were conducted. Visual Analogue Scale (VAS) and Standard Gamble (SG) were used to reveal preferences on a scale running from 1 (perfect health) to 0 (all worst or death depending on the interviewee's decision). The 8 single-attribute value functions were directly revealed as well as the corner states values. Rating of particular multi-attribute health states was also done using both VAS and SG. **Results :** Firstly, our survey demonstrates, for the 1st time, the feasibility of eliciting values and utilities in France. Secondly, analysis concludes that collected data permit 1) to exhibit the importance to take into account individual's risk attitude when modelling preferences for health states 2) to fit multi-attribute utility functions for many individuals and 2) to discuss the appropriateness of the multiplicative functional form. Results obtained at the person-mean and person-median levels will be presented and discussed as well as the strong concordance between observed and estimated utilities.

Abstract 1658**EFFECTS OF DEMOGRAPHIC FACTORS ON UTILITY SCORES FOR HYPOTHETICAL HEALTH STATES**

William J. Furlong, Charles H. Goldsmith, CE&B, McMaster University, Hamilton, ON, Canada

General population utility scores are being used to evaluate health services. A widely held hypothesis is that these scores vary by demographic factors. This study assessed score variability explainable by a variety of demographic factors: age, education, gender, income, marital status, native language, number of children, occupation, and religious participation. Utility scores and demographic data were from 4 surveys in Hamilton ON, Canada of adult general population random samples conducted for 3 studies: Child Health Index Study (64 raters of 6 states); Childhood Cancer Study (203 raters of 4 states); and Prenatal Diagnosis Health-Related Quality of Life Study (64 female raters of 5 states and 64 male raters of 5 states). Trained professional interviewers collected time trade-off or standard gamble scores in highly-structured interviews. Variability in scores was assessed by ANOVA and step-wise regression. Size and direction of differences in mean scores for groups were examined. Predictor variable collinearity and multiple-testing issues were addressed. Single-variable analyses for all 4 surveys showed statistically significant score variation due to 2 factors: health state descriptions and individual raters. Approximately 2/3 of total variance was explained by these 2-factor models. Unexplained variability was remarkably similar across surveys: 35.7% to 37.9%. Inter-rater variability ranged from 32.8% to 55.6%. Age, income, number of children and occupation were significant ($p < 0.005$) in at least 1 survey. Age, income and number of children were each significant factors in 2 surveys, however the direction of differences was inconsistent. No factors were significant in more than 2 surveys. A small percent of inter-rater variability is explained by demographic factors. Statistically significant factors were associated with small differences in mean scores. In sum, the 5 commonly measured demographic factors (age, education, gender, income, marital status) did not explain important amounts of the variability in utility scores.

Abstract 1564**HOW GENERALIZABLE ARE MULTI-ATTRIBUTE UTILITY FUNCTIONS?**

David H. Feeny, Qinan Wang, William Furlong, George W. Torrance, Ronald D. Barr, Institute of Health Economics, Edmonton, AB, Canada

How Generalizable are Multi-Attribute Utility Functions? D Feeny, Q Wang, W Furlong, G Torrance, and R. Barr Institute of Health Economics, Edmonton, Alberta, Canada **Purpose.** The purpose of the study was to compare a multi-attribute utility function based on scores obtained from a random sample (n = 194) of general population (GP) parents with a function based on scores from a sample (n = 59) of parents of patients (PP) with cancer. It was hypothesized that the PP sample might have importantly different preferences. **Methods.** A random sample of GP parents was drawn, along with a parallel sample of parents of children on treatment for 3 cancers. The same trained professional interviewers conducted both surveys. Feeling Thermometer (FT) scores were obtained for 34 Health Utilities Index Mark 2 (HUI2) health states; standard gamble (SG) utility scores were obtained for 4. Power functions to convert FT to SG scores were estimated. Scores were compared using 2-way analysis of variance (ANOVA). Parameter values for utility functions were compared. Scores for 14 states derived using the two functions were compared. **Results.** The response rates were 71% for GP parents and 92% for PP. FT scores for 16 single-attribute levels were similar (mean difference per state (MD) = 0.02). There were statistically significant differences only for emotion. FT scores for 12 states were similar (MD = -0.01). SG scores for 4 states were similar (MD = -0.01). Exponents for the power functions were 2.29 and 2.16. The 8 parameter values for the utility functions were similar (MD = 0.005). The two utility functions performed similarly in predicting the 4 validation states with MD per state of 0.04 (GP) and 0.05 (PP). Scores calculated for 14 states observed among survivors of childhood cancer were similar (MD = 0.01). A utility function based on scores from PP was very similar and almost interchangeable with one based on scores from GP parents. In this case the functions are highly generalizable.

Abstract 1010**THE EFFECTS OF HEALTH-RELATED QUALITY OF LIFE (HRQOL) MEASURED BY THE SF-36 ON CONSULTATION RATES OF A CHINESE POPULATION**

Cindy L. Lam, Ian J. Lauder, Tai Pong D. Lam, Family Medicine Unit, Department of Medicine, The University of Hong Kong, Hong Kong SAR, China

Clinicians and health administrators often question the clinical significance of HRQOL and wonder how its assessment can benefit their work. Studies in Western populations have shown that health status is an important determinant of service utilisation but an earlier study on a group of Chinese elderly failed to show any correlation. The aim of this study was to find out whether HRQOL measured by the SF-36 was an important independent determinant of consultation rates for the general Chinese population in Hong Kong. This was a cross-sectional telephone survey of 2410 Chinese adults randomly selected from the general population in Hong Kong. The subjects aged 18 to 88 (mean 42.9) years old, 52.2% of them were females and 38% had one or more chronic diseases. The Chinese (HK) SF-36 and a structured questionnaire on demographic characteristics, chronic morbidities and annual and monthly consultation rates were administered to each respondent by telephone. The annual and monthly consultation rates were regressed on demographic factors, morbidity variables and the SF-36 scores by multivariate stepwise linear regression. The regression model explained 27.5% and 21.3% of the total variance in annual and monthly consultation rates, respectively. The SF-36 scores were stronger determinants of consultation rates than demographic or morbidity factors. They nearly doubled the predictive power for annual consultation rate and tripled that for monthly consultation rate when they were added to the model with demographic and morbidity variables only. This was the first population-based study to show that HRQOL had a significant impact on the health service utilisation of the Chinese. The SF-36 scores were important independent determinants of consultation rates and they could improve the predictive power of medical risk-adjustment models.

Abstract 1368**INDICATORS OF IMPAIRMENT AND PERCEPTION OF IMPAIRMENT IN TREMOR ARE RELATED THROUGH A NON-LINEAR RELATION**

Michael A. Hagan, Steven M. Albert, Elan D. Louis, Division of Sociomedical Sciences, Columbia University, NYC, NY

The relationship between QoL indicators of impairment (Iol) reported by an objective medical observer (Independent Rating of Performance) and perception of impairment (Pol) as indicated by a patient's self-rating of disability caused by tremor (Self Report of Tremor) was examined in a prospective cohort study. Data from 113 subjects enrolled in the Functional Assessment of Tremor study, were analyzed through regression analysis where Iol and Pol, as the independent and dependent variables respectively, were collected on the same functional indicators. Analyses progressed from fitting linear to non-linear monotone and finally non-linear non-monotone functions to determine the functional form between Pol and Iol. The squared correlation was only 0.437 for the linear data showing that a simple linear model is not appropriate for these data. By relaxing the constraints placed on the regression line, the proportion of the variance accounted for increased from 0.437 (linear) to 0.647 (non-linear monotone) to 0.667 (non-linear non-monotone). The last two figures suggest that the relationship between Pol and Iol increases monotonically and it is not cyclical since a non-linear, non-monotone model increased the squared correlation by only 0.02 units. Fitting a piecewise-linear-regression confirmed these results. Although findings indicate that patients assess their condition well, results suggest a perceptual threshold for self-reported disability towards the top of the curve where the slope flattens out. This should be expected if a compensatory mechanism helps subjects adapt to disability, such that their perception of their own disability diverges from that of an independent/objective observer. The flattening of the slope may also be due to a ceiling effect. The data is not sufficiently robust to differentiate between a quadratic or cubic relationship but clearly establishes a non-linear relationship. Funding: Beeson Award to Elan Louis, MD

Abstract 1656**ARE CONDITION-SPECIFIC MEASURES TRULY CONDITION-SPECIFIC?**

Elaine McColl, CHSR, University of Newcastle, Newcastle upon Tyne, UK

Condition-specific instruments, it is argued, should measure only the impact of the index condition and associated interventions. But respondents' comments suggest that some have difficulty in separating the effects of index and co-morbid conditions. We therefore investigated whether there was a significant difference in scores on both generic and condition-specific quality of life measures between respondents who had a range of co-morbid conditions and those who did not. The subjects were adults with asthma or angina, drawn from 60 family practices in north-east England. The generic instruments were the SF-36 and EQ-5D; the condition-specific measures were the Seattle Angina Questionnaire (SAQ), the Asthma Quality of Life Questionnaire (AQLQ) and an asthma-specific measure of symptom frequency. All respondents indicated whether they suffered from specified chronic co-morbid conditions. For the purposes of analysis, these were grouped according to the 'system' affected (eg 'respiratory condition', 'heart disease'). To test whether the presence of a co-morbid condition influenced quality of life scores, mean differences (presence - absence) and associated 95% confidence intervals were calculated. Since both co-morbidity and quality of life are significantly associated with gender and age, regression models were used to test if any effects of co-morbidity remained after controlling for these factors. For both angina and asthma, patients with co-morbid conditions reported poorer quality of life on the generic measures. However, significantly lower scores were also observed on the condition-specific measures. For example, patients with arthritis and/or multiple sclerosis, had scores on the Physical Limitations subscale of the SAQ that were on average 11.3 points lower than patients who did not have a mobility-limiting condition. Observed differences in condition-specific quality of life scores persisted after controlling for the effects of age and gender. We conclude that respondents may find it difficult to 'isolate' the impact of the index condition.

Abstract 1698**ARE PATIENT-BASED MEASURES SUSCEPTIBLE TO THE PLACEBO EFFECT?**

Bruce Crawford, Stanley Cohen, Joseph Smollen, Peter Tugwell, Vibeke Strand, MAPI Values, Boston, MA

A pronounced placebo effect is common in clinical efficacy measures (eg., tender joint count [TJC], swollen joint count [SJC], and physician global assessment [PHGA]) in rheumatoid arthritis (RA) trials. Patient-based measures of function are part of the ACR response criteria to assess drug treatment. Placebo data from the US301 trial which compared leflunomide to methotrexate and placebo were analyzed (see table) as % improvement from baseline to assess placebo effect for physician-based measures (TJC, SJC, PHGA) and patient-based measures (the Modified HAQ [MHAQ], HAQ Disability Index [HAQ DI], the Problem Elicitation Technique weighted Top 5 [PET], SF-36, and Patients Global Assessment [PGA]). Results showed a typical placebo effect in TJC, SJC, and PHGA, but not in MHAQ, HAQ DI, PET, SF-36, or PGA. We conclude that patient-based measures of drug effect are less susceptible to the placebo effect seen with physician-based efficacy measures in RA trials.

Placebo Data (n=118)		6 Months	12 Months
Measure	BL mean	% improved (% worsened)	% improved (% worsened)
Tender Joint Count	16.5	18.4	15.6
Swollen Joint Count	14.8	21.5	19.9
Physician Global Assessment	6.2	19.9	16.1
MHAQ	0.89	(12.1)	(15.7)
HAQ DI	1.31	(3.9)	(4.8)
PET Weighted Top 5	22.4	(17.2)	(22.1)
SF-36 PCS	28.9	6.5	6.3
Patients Global Assessment	5.8	(4.3)	(14.1)

Abstract 1750**A COMPARISON OF THERAPIST EVALUATION AND CLIENT SELF-EVALUATION USING THE SF-36**

Mary E. Buggy, Christine Flack, Steve Shapiro, Jerry Hagene, Barbara Gandek, George P. Sillup, Counseling Center, Daemion House, Berwyn, PA

The purpose of this research was to evaluate the use of a health-related quality of life instrument as an aid to therapists' assessments of clients' progress at intake and selected intervals during ongoing counseling or completion of therapy. Sixteen subjects (14 females, 2 males) were randomly selected from the 148-client cases at Daemion House, a counseling center serving the working poor. Client diagnoses ranged from family dysfunction to bipolar disorder, depression and combined medical problems. Therapists conducted and documented client interviews at intake (baseline), after regularly scheduled counseling sessions or at completion of therapy for an average period of 1 year. During that same time at baseline and the most recent/final session, clients were asked to complete the MOS Short Form (SF-36). Therapists were blinded to the SF-36 scores, which were tabulated by a computer-based system that protected clients' identities and recorded/cross-checked domain scores. A third party psychologist statistically compared SF-36 scores to evaluate clients' self-assessments and qualitatively compared the SF-36 results with the therapists' assessments. SF-36 results were congruent with therapists' assessments for 11 of the 16 clients (5 family dysfunction, 4 depression, 1 bipolar disorder, 1 combined medical problems). Ten of 11 (8F, 2M) showed improvement consistently in 4 domains, Mental Health (MH), Social Functioning (SF), Role Emotional (RE) and Vitality (VT) (p<.05). One did not show improvement and had correspondingly negative MH, SF and VT scores (p<.05). Five clients had therapist evaluations that were different from the SF-36 scores; 4 had scores in 7 of the domains that were consistent with the therapists' evaluations except for MH, which was consistently different (p < .05). The SF-36 is a useful adjunct to therapists' evaluations, especially to highlight changes in domains related to emotional stability.

Abstract 1160**CONCOMITANT RADIOCHEMOTHERAPY FOR OESOPHAGEAL CARCINOMA (FNCLCC-FFCD 9305): SPLIT-COURSE RADIOTHERAPY DOES NOT IMPROVE PATIENTS' QUALITY OF LIFE.**

Jean-Paul Moatti, Pascal Auquier, Geneviève Macquart-Moulin, Jean-François Seitz, Michel Henry-Amar, Christiane Langlois, Olivier Bouché, Jacques-Henri Jacob, INSERM U379, Marseille, France

Withdrawn

Abstract 1268**HEALTH-RELATED QUALITY OF LIFE IN LONG-TERM HEAD AND NECK CANCER SURVIVORS: A COMPARISON WITH GENERAL POPULATION NORMS**

Eva Hammerlid, Charles Taft, Marianne Sullivan, Department of Otolaryngology Head and Neck Surgery, Sahlgrenska University Hospital, Göteborg, Sweden

Purpose: To examine the health-related quality of life (HRQL) in long-term head and neck (H&N) cancer survivors compared with general population norms. HRQL was assessed with three standardised questionnaires: the SF-36 Health Survey (Short Form 36) and the EORTC QLQ-C30 and QLQ-H&N35 (European Organization for Research and Treatment of Cancer Quality of Life Questionnaire, -Core 30 and H&N35 cancer module). Altogether 135 H&N cancer patients (mean age 62 years, 31% females) of 151 survivors (89% acceptance) from a longitudinal HRQL study (n=232) were included three years after diagnosis. **Results:** The H&N cancer patients' SF-36 scores did not differ significantly from those of an age- and sex-matched sample (n=871) from the Swedish normative population, except on the role-physical functioning scale. On the other hand, treatment-related side-effects and disease specific problems (e.g., swallowing, local pain and dry mouth) measured by the H&N cancer module were, with few exceptions, significantly worse than norm values. Gender comparisons revealed that female H&N cancer patients generally scored better than the norms on both the SF-36 and the EORTC QLQ-C30, while the male patients scored significantly worse on most SF-36 scales. Patients >65 years more often scored worse than the norm than did patients <65. Clinically relevant differences were found on the majority of SF-36 scales in comparison of tumour sites, however, comparisons of patients with small (stage I+II) versus advanced (stage III+IV) tumours revealed few differences. **Conclusion:** Three years after diagnosis H&N cancer patients still suffer significant functional limitations/problems related to their disease and its treatment but these problems do not generally affect their overall HRQL. Tumour stage no longer differentiates HRQL at three years, however, factors related to the patients' age, gender and location of the tumour appear to have bearing on their reported health status.

Abstract 1722**QUALITY OF LIFE OF LONG-TERM SURVIVORS OF HEAD AND NECK OR PROSTATE CANCER: PRELIMINARY RESULTS FROM A SURVEY OF SURVIVORS FROM RADIATION THERAPY ONCOLOGY GROUP (RTOG) CLINICAL TRIALS.**

Charles B. Scott, Jean Stern, Sucha O. Asbell, Harmar D. Brereton, Jan Peer, Phillip Littman, Charles Scarantino, Deborah Bruner, Department of Quality of Life Research, RTOG, Philadelphia, PA

This research project was designed to evaluate the QOL of prostate cancer survivors (PCS) or head and neck cancer survivors (HNCS) enrolled on RTOG clinical trials. Patients alive > 4 years from registration on RTOG clinical trials were eligible to participate. Potential PCS or HNCS were identified in the RTOG database and institutions (INST) that agreed to participate were sent surveys and a list of eligible survivors. All eligible PCS or HNCS at that INST were given an informed consent and a survey. The survey consists of questionnaires on QOL, insurance issues, mood, sexual function, alcohol and tobacco use, and mental status. Demographics and mood results are presented here. To date, 447 survivors were approached from 40 INST and 240 (54%) have signed the informed consent. The primary reason for non-participation has been patient refusal. Twenty-one percent are HNCS. Sixteen percent of PCS are African American, as are 10% in HNCS. The current average age of PCS is 76 (range of 55-91 years); 66 (41-84) for HNCS. Current smokers comprise the minority of the sample: 8% in PCS, 13% in HNCS. Alcohol use was measured as number days having a drink per month; its use is greater in HNCS average 16 days/month vs. 11 days in PCS. The majority of HNCS have a normal diet (72%) and understandable speech (74%). The Profile of Mood States was used to evaluate mood. In comparison to normative values, both PCS and HNCS had less anxiety, depression, anger and fatigue, but confusion was increased. Preliminary results of this survey indicate that long-term survivors have excellent functioning and positive outlook. This sample is generally retired and elderly, but not fatigued. However, as expected with a geriatric sample cognitive functioning issues are dominant. An evaluation of mental status results needs to be performed.

Abstract 1749**HEALTH-RELATED QUALITY OF LIFE OF PATIENTS WITH PROSTATE CANCER: LOCALIZED VERSUS ADVANCED DISEASE**

JB Madalinska, ML Essink-Bot, H J. de Koning, W J. Kirkels, P J. van der Maas, F H. Schroder, Department of Public Health, Erasmus University Rotterdam, Rotterdam, The Netherlands

The current study was undertaken within the framework of the European Randomized Study of Screening for Prostate Cancer (Rotterdam screening trial) to compare the health-related quality of life (HRQOL) effects of localized and metastatic prostate cancer. We conducted a cross-sectional study among patients with localized and advanced prostate cancer using both generic and disease-specific HRQOL measures (SF-36, EuroQoL, QLQ-C30, UCLA Prostate Cancer Index). Clinical data were retrieved from patient medical records. For the patients with advanced prostate cancer, a 24-month follow-up investigation was performed to determine mortality and progression/remission rates. Data were available on 82 patients with M1 disease and on 100 patients with localized disease (stages: T1-T3) at 3 months after primary treatment (response rates: 62-93%). Patients with advanced disease were significantly older than those with localized prostate cancer (66 vs. 73 yrs., $p < .01$). Cross-sectional analyses (adjusted for the effect of age) revealed significantly poorer levels of generic HRQOL (the SF-36 and EuroQoL scores) in patients with advanced prostate cancer, compared to men with the localized disease, although both groups scored below the population norm. Patients with advanced disease reported significantly more problems with pain, fatigue, urinary and bowel functioning compared to patients with the localized disease ($p < .05$). Additionally, 40-60% of the patients from the first group reported side effects from hormonal treatment. At two years after the HRQOL assessment, 61% of patients with metastases were alive (remission: 33%, progression: 28%). HRQOL of patients with advanced prostate cancer is severely impaired. If screening reduces mortality from prostate cancer, it will also prevent metastatic prostate cancer. At the population level, the decrease in life-years lived with poor HRQOL associated with advanced disease is favorable effect of screening.

Abstract 1114**HEALTH OUTCOMES AFTER RADICAL PROSTATECTOMY OR RADIOTHERAPY FOR CLINICALLY LOCALIZED PROSTATE CANCER**

Arnold L. Potosky, Division of Cancer Control and Population Sciences, National Cancer Institute, Bethesda, MD

Radical prostatectomy (RP) and external beam radiotherapy (RT) are the two major therapeutic options for treating men with clinically localized prostate cancer. Because survival is often favorable, treatment decisions may depend on the probability of long-term complications. Patients diagnosed with clinically localized prostate cancer between the ages of 55-74 who received either radical prostatectomy (n=1,156) or radiotherapy (n=435) were included in this population-based study. Medical record abstraction and patient surveys were used to ascertain baseline characteristics and to track changes in outcomes. Men receiving radical prostatectomy (RP) were more likely than those receiving radiotherapy (RT) to be incontinent almost 2 years after treatments (9.6% vs 3.5%, p<.001). Among the RP patients, there was an acute increase in incontinence within the first 4 months following treatment, followed by a recovery period which continued through the second year. RP patients reported less diarrhea at 2 years (20.9% vs 37.2%, p<.001), and a similar pattern was also observed for bowel urgency. Larger decrements in bowel function were observed in men receiving RT. RP patients had much higher rates of impotence at 2 years (79.6% vs 61.5%, p<.001). These differences persisted after adjustments for propensity score. RP patients experienced larger declines in urinary function, while RT patients experienced greater declines in bowel function. Large, significant declines in sexual function were observed in both treatment groups. The treatment groups were similar in terms of general dimensions of HRQOL. In contrast to earlier findings in smaller, selected patient samples, these outcome differences reflect treatment delivered to a heterogeneous group of patients in all types of health care settings. Our results provide comprehensive and representative information about long-term complications to help guide and inform treatment decisions.

Abstract 1363**THE IMPROVEMENT OF QOL-RELATED AREAS IN HEALTHY, COGNITIVELY INTACT ELDERLY PERSONS BY THE GINKGO BILOBA EXTRACT EGB 761**

A. Cieza, F. Rösch, H. Hägele-Kaddour, E. Pöppel, Institute for Medical Psychology, Munich, Germany

During the past 15 years, many studies have described the positive effects of EGb 761 (Tebonin®, Dr. Willmar Schwabe Pharmaceuticals, Karlsruhe, Germany) on emotional and mental functioning, especially in patients with cerebral disorders of vascular origin, Alzheimer-type dementia, or depression. Nevertheless, there is a lack of investigations examining the effects of EGb 761 on healthy subjects. One purpose of this study was to evaluate the effects of EGb 761 on some areas related to QoL, like self perception of well-being, emotional and mental states in healthy, elderly volunteers without age-associated cognitive impairment. The trial was conducted as a comparative, parallel-group, double-blind trial. 66 subjects aged between 50 and 65 years were randomized, 32 into the placebo and 34 into the EGb761-treatment group (240mg, tid). At Baseline and after 28 days of treatment, the subjects completed the following QoL questionnaires and scales: Profile of Mood States (POMS), the Self Rating Depression Scale (SDS), the VAS-QoL, the VAS-General Health, the VAS-Mental health, and the Subjective Intensity Scale-Mood (SIS-Mood). The test statistics of the parametric test (t-test) were adjusted on the basis of the estimation of the variability of the measurement error for each analyzed variable to compare overall intergroup and intragroup results of the two trials populations. Significant improvements were found within the verum group but not within the placebo-group when the intragroup results at Baseline and Final Examination were compared in the following variables: POMS-Depression, SDS, VAS-MH and SIS-Mood. Intergroup differences were also found at the final Examination for the VAS-MH. In addition to the significant improvement on action/reaction and activation/attention observed within this study, it can be concluded that the intake of EGb 761 is associated with a more positive self perception in some components related to QoL, like mental health and mood state in healthy volunteers.

Abstract 1101**THE PROFILE OF SF-36 IN A COMMUNITY-DWELLING ELDERLY SAMPLE: HIGH VS LOW FUNCTIONED FEMALE ELDERLY**

Beeto, Wai-Chung Leung, Department of Psychology, The Chinese University of Hong Kong, Hong Kong, China

The SF-36 has been used as a generic measure of health-related quality of life in general populations and in a variety of patient groups. However, few studies have focused on investigating the profile of healthy elderly who are without any disease or impairment. The purpose of the present preliminary study was to compare the profile of SF-36 between healthy elderly with high and low functioning. Forty-eight female participants aged from 65 to 79 (mean age = 72.25; S.D.=4.21) was recruited in a community elderly center. Face-to-face interviews were conducted and SF-36, MMSE and IADL were administered as part of a comprehensive battery. Those who scored in the upper quartile in their age- and education- adjusted MMSE and reported no disability in IADL were defined as high functioning while others, not meeting the above criteria, were defined as low functioning. Significant difference were found between RP, RE and MH scores which indicated that elderly with high functioning tended to have less role limitations in carrying out daily activities due to their physical and emotional problems and they tended to have better mental health. Also, elderly defined as high functioning seemed to have better physical functioning, less bodily pain and better social functioning measured by PF, BP and SF respectively although the differences were not significant. In general, elderly defined as high functioning (by MMSE & IADL) tended to have better score in SF-36. Individual scale scores of SF-36 may be useful in predicting the functioning level of healthy elderly. Further studies on the effect of aging on SF-36 were suggested. (Scale scores with * indicate a sign. difference; p<.05)

Mean (SD) of SF-36 scale scores: Low & High functioning group

	PF	*RP	BP	GH	VT	SF	*RE	*MH
Low	60.14 (16.65)	29.29 (34.02)	58.42 (27.03)	53.14 (23.22)	57.50 (21.50)	66.43 (22.64)	45.37 (42.)	64.56 (19.53)
High	70.00 (22.96)	60.14 (16.65)	60.14 (45.80)	53.25 (19.11)	54.17 (28.11)	73.96 (27.93)	75.00 (35.17)	75.33 (13.42)

Abstract 1714**CROSS-CULTURAL COMPARISON OF OLDER ADULTS WITH MODERATE DISABILITIES: AN APPLICATION OF Q-METHODOLOGY**

Judy Kruger, Carsten Konig, School of Public Health, University of Illinois at Chicago, Chicago, IL

Context: There is an increased tendency for health professionals to use survey research methods to understand the meaning of qol for older adults, however, often the instruments are ill suited to life issues relevant to older adults. Subjective methods may be more appropriate to understanding this heterogeneous population and can build on the current body of knowledge. Purpose: The aim of this study was to explore a method for achieving consensus about qol for older adults. Subject sample: Subjects were selected from congregate communities in Chicago, USA and in Duesseldorf, Germany. All of the participants were relatively healthy and had minor adl limitations. Method: The method used was a 'nominal' group approach to focus attention on the complexities and conceptualizations with approximately 40 participants who were asked to prepare q-sorts, a form of preferential ranking and were analyzed by-person factor analyses. The application of Q-methodology allows cultural comparisons using older adults' choice of words and adds conceptual clarity. Results: Three factors were found to be relevant to older adults, which accounts for different viewpoints labeled contentment, community connection and faith. This study was designed to further understand and develop the concept of qol with a particular focus on older adults' perceptions of the meaning of qol.

Abstract 1289**QUALITY OF LIFE IN CATARACTOUS PATIENTS TREATED WITH DAY SURGERY**

Takashi Mandai, Kozaburo Adachi, Kanehisa Morimoto, Shunji Tsuboi, Haruyasu Fujita, Soichiro Maruyama, Fumio Nomura, Yoichi Kakuta, Takuya Yoshida, Sachiyo Okabayashi, Tomoko Namba, Ken Matsuoka, Japanese Society of Quality of Life Research, Kobe, Hyogo, Japan

The purpose of this study was to evaluate the quality of Life(QOL) in cataractous patients treated with day surgery. Seventy-four cataractous patients treated with day surgery participated in this study. Both personal interviews and our self-administered questionnaire including 40 questions divided into 15 categories with the Life Satisfaction Index were used. Pearson's correlation coefficients between our questionnaire and the Life Satisfaction Index was $r=0.92(P<0.05)$. Cronbach's alpha coefficients of our questionnaire were high enough to accept for clinical use; 0.93 in dietary problems, 0.91 in mental function, 0.88 in economical problems, 0.86 in well-being and work performance etc, respectively. Our questionnaire contained 11 main factors and cumulative contribution was 0.82. After day surgery, one half patients indicated the improved total QOLs but another one showed the deteriorated total QOLs. In the improved QOL patients, significant improvements of QOL were demonstrated in the strata of medical service, physical function, well-being and social participation ($p<0.05$), but there were significant deteriorations of QOL in mental function i.e. concentration and memory($P<0.05$) in the deteriorated patients, too. These findings indicate that our questionnaire has high enough reliability and potency of validity to use for cataractous patients. Although we recognize the merit and demerit of day surgery, we are easy to understand the merit of no necessity of admission for keeping the high QOL in perioperative period. We should pay more attention for the lack of mental postoperative care and counseling in the complicated cases.

Abstract 1483**ARE THE DAY CENTRES DIFFERENT FROM HOMES IN WHAT CONCERNS ELDERLY FUNCTIONAL STATUS?**

Pedro L. Ferreira, Rogério C. Rodrigues, School of Economics, University of Coimbra, Coimbra, Portugal

During the second half of the twentieth century a new phenomenon emerges: the demographic aging characterized by a low fertility rate, a low mortality rate, reduction of the household, and population increase close to zero. This has social and health implications, namely the increase of needed care, growth of the degenerative diseases and higher vulnerability. Multiple pathology, complications, chronicity, sensorial and movement limitations induce social handicaps and the need for healthcare planning. The OARS (Older American Resources and Services) is one of the best available condition specific measurement instruments for this age group. It is composed by two distinct parts. Part A deals with functional assessment in five areas (social resources, economic resources, mental health, physical health and activities of daily living); part B assesses the use and the perceived need of six service groups (general support, social and recreative services, health services, economic support services, support, assessment and coordination services and non classified services). The objectives of this study include comparing the users of day centres with the users of elderly homes in what concerns functional disability prevalence, distribution of individuals with similar care needs, the use of services and their perceived need. The study population included 20,689 elderly (older than 65 years), 10,644 being day centre users; the probabilistic sample obtained was formed by 155 day centre users and 147 home residents from 24 institutions. In general, comparing to day centres, in elderly homes we evidenced a lower functional status, lower income and economic resources, lesser number of consultations, lesser social resources, higher disabilities and lesser activities of daily living.

Abstract 1089**BUT WHAT IS A LOGIT, ANYWAY? - INTERPRETATION OF QUANTITIES IN IRT METHODS**

Rebecca Holman, Robert Lindeboom, Clinical Epidemiology and Biostatistics, Academic Medical Center, Amsterdam, The Netherlands

Following dissatisfaction with classical methods, IRT continues to increase in popularity. It has been acknowledged that the use of this method in conjunction with multi-item instruments designed to measure patients' functional status or level of quality of life has a number of advantages over total score methods. These include measurement properties such as the linearity of the measures of QoL, and practical advantages such the use of (computer aided) adaptive testing. However a major disadvantage and disincenitive for the use of these techniques is the lack of clarity about the meaning of quantities expressed in logits. Many experienced researchers feel that they have developed an intuitive idea of the meaning of a point on a classically scored instrument, which they know well. In contrast, logits can appear to be more distant from the instrument and the population under consideration. This presentation will use graphical methods to relate these quantities to the subject matter of the questionnaire used and the patient population under investigation. This means that the logits can be clearly compared with familiar concepts. These methods will also be used to enhance visual presentation of IRT scales, and understanding of treatment effects in trials. Firstly the results of a preliminary calibration of 75 items from the ADLs item bank, designed to measure ADL functioning, on approximately 400 outpatients in our hospital will be used as a general example. Secondly, the data from a recently published clinical trial into the efficacy of surgical treatment options for essential tremor will be used as an illustration of the power of these methods in a clinical trial. It can be concluded that it is possible to gain a practical understanding of quantities expressed in logits using one of the methods described above. In addition it is hoped that the application of these simple techniques will enable many more investigators to appreciate the advantages of IRT techniques.

Abstract 1276**WHAT AFFECTS THE INTRUSIVENESS PATIENTS EXPERIENCE DUE TO THEIR DISEASE?**

Yasmin Maor, Liraz Olmer, Benjamin Mozes, The Center for the Study of Clinical Reasoning, The Gertner Institute, Tel Hashomer, Israel

Patients may be affected in many ways by their illness. One of the aspects of disease impact on patients is the amount of intrusiveness they suffer in daily life. That is, the intensity of subjective stress a patient perceives due to his/her illness. This approach regarding disease burden was not investigated. Therefore, our aim was to investigate what variables affect the intrusiveness patients experience due to their illness. 219 patients with chronic lung disease underwent a structured interview. In addition, spirometry, Hb-S% at rest and during exercise and the distance walked during 6 minutes (6MWD) were measured. The Disease Intrusiveness Measure (DIM) used was a modification of the intrusion part of the Impact of Events Scale, which includes three questions regarding the intrusiveness of illness. Questions relate to involuntary time spent thinking about the illness or symptoms, visions regarding the illness or symptoms, and sleeping problems due to thoughts or visions regarding the illness or symptoms. The scale ranges from 3-15. Hierarchical multivariate analysis was employed to find relations between candidate variables and the DIM. Candidate variables were sociodemographic details, biologic parameters, symptoms, objective physical performance, perceived physical and social functioning, and depressive symptoms and anxiety. Mean age was 55 years (SD 18 years). 63% were males. 36% had chronic obstructive lung disease. Mean percent of observed divided by predicted forced expiratory volume in the first second (FEV₁) was 60% (SD 23%). Overall we explained 44.9% of the variance of the DIM. While severity of physical symptoms (dyspnea and pain) explained 20.7% of the variance, mental symptoms explained another 18.9%. In this population of chronic lung patients, the intrusiveness patients experience due to their disease is highly affected by physical symptoms and emotional state.

Abstract 1440**GUIDELINES FOR INTERPRETATION OF THE GRAVES' OPTHALMOPATHY QUALITY OF LIFE QUESTIONNAIRE (GO-QOL)**

Caroline B. Terwee, Friedo W. Dekker, Mark F. Prummel, Wilmar M. Wiersinga, Department of Clinical Epidemiology & Biostatistics, Academic Medical Center, Amsterdam, The Netherlands

The GO-QOL is the first instrument available to measure Health-Related Quality of Life (HRQL) of patients with Graves' ophthalmopathy. The objective of this study was to test its longitudinal validity and to provide guidelines for defining a minimal clinically important difference (MCID) in score on the GO-QOL that can be considered an important improvement in HRQL for patients. The MCID can be used for interpretation of study results and in sample size calculations. We included 164 patients scheduled for 5 different treatments. Patients completed the GO-QOL (2 subscales), SF-36, SIP and EuroQol, before and 3 or 6 months after treatment. Mean changes after treatment in the most relevant GO-QOL subscale were 8-20 points after major treatment (radiotherapy or decompression) and 3-4 points after minor surgery (eye muscle surgery, eyelid lengthening, blepharoplasty). A clinical response to treatment was associated with a change in GO-QOL scores of 10-20 points after major treatments and 3-10 points after minor surgery. Changes of about 6 points were already considered important improvements by the patients themselves. These data provide benchmarks for the interpretation of GO-QOL changes in future studies. The direction and amount of change in GO-QOL scores after different treatments were in accordance with our prespecified hypotheses about treatment effects. Effect sizes (ES) in the GO-QOL subscales were generally higher than ESs of SF-36 subscales, supporting the longitudinal validity of the GO-QOL. We argue that the MCID cannot be considered a fixed property of an instrument because aspects like placebo effects, treatment burden, and heterogeneity of the patient population should be taken into account. As a general guideline, we recommend a mean change of at least 6 points an important change in HRQL for patients. For more invasive therapies, a change of at least 10 points is recommended for sample size calculations.

Abstract 1770**GOAL INTERFERENCE AS A CRITERION FOR THE MEANINGFULNESS OF QUALITY OF LIFE SCORES**

Amy H. Peterman, Josephine Ribaldo, Nicole Laliberte, Marianne J. Brady, David Cella, Robert H. Lurie Cancer Center, Northwestern University Medical School, Evanston, IL

There are now many reproducible and valid QOL questionnaires available for clinical trials and clinical practice evaluation. However, our understanding of the personal meaningfulness of scores on these instruments remains limited and a criterion that has individual relevance would be a useful tool. The Goal Interference Scale (Peterman, et al., 2000) measures the degree to which disease/treatment interfere with a person's ability to make progress on personally meaningful projects. When QOL scores are mapped to the extent of goal interference, it may aid our interpretation of the difference in magnitude between QOL scores. This was examined in a sample of 32 subjects of mixed cancer diagnosis who were currently receiving chemotherapy; the sample was 75% female and the median age was 61 years. The Functional Assessment of Cancer Therapy-General (FACT-G, Cella et al., 1993) and the Goal Interference Scale (Peterman, et al., 2000) were administered: after nominating current important personal goals/projects, subjects were asked about the degree to which "the effects of disease and/or treatment interfere with your progress on this project." Three groups were created reflecting those with high interference (n=9), medium interference (n=10), and low interference (n=13) scores. The cross-sectional relationship between QOL scores and the degree of goal interference was examined within a generalized linear model. Physical well-being [F(2,30)=3.6,p=.04], functional well-being [F(2,29)=3.4,p<.05], and total FACT-G [F(2,29)=3.8,p<.05] scores differed significantly between the three groups. Goal interference group accounted for approximately 20% of the variance in QOL scores (R²s=.19 - .20). The degree of goal interference created by disease/treatment may provide a useful method for evaluating the clinical meaningfulness of QOL scores.

Abstract 1747**VALIDATION OF THE HEADACHE IMPACT TEST IN RELATION TO HEADACHE PAIN SEVERITY AND PROBABILITY OF MIGRAINE DIAGNOSIS**

John E. Ware, Jr., J Bjerne, M Kosinski, M Diamond, MB Bayliss, S Tepper, A Dowson, AS Batenhorst, QualityMetric, Inc., Lincoln, RI

Purpose: To test the validity of the Headache Impact Test (HIT) in relation to migraine diagnosis and headache severity. Sample and Methods: Item response theory (IRT) methods were used to calibrate all 54 items from the Headache Disability Index (HDI), Headache Impact Questionnaire (HIImQ), Migraine Disability Assessment (MIDAS) and Migraine-Specific Quality of Life Questionnaire (MSQ). HIT used IRT parameters and computerized adaptive methods to estimate impact scores from a small subset of items. Survey data from 1,016 representative headache sufferers who completed all items were analyzed. International Headache Society (IHS) criteria were used to identify migraine sufferers (n=746) and previously validated criteria were used to stage severity. Original instruments were scored using developers' algorithms. Two IRT-based summary scores were estimated: HIT-Total (53 items) and HIT-Dynamic (5 items or fewer using a simulated dynamic administration). Analyses of variance (ANOVA) and relative validity (RV) tests were performed to compare all measures in terms of their validity in discriminating headache pain severity and diagnosis. Summary of Results: All instruments significantly discriminated levels of severity and migraine diagnosis. F-statistics ranged from 540.6 to 145.2 for migraine diagnosis and 49.0 to 222.1 for headache severity. HIT Total discriminated best for both diagnosis and severity; RVs were 99% and 98% for HIT Dynamic in these tests. Conclusions: Dynamic assessments of headache impact may achieve the brevity of a short form while approximating the empirical validity of a long form.

Abstract 1468**A NEW METHOD TO ASSESS THE EQUIVALENCE OF QUALITY OF LIFE (QOL) SCALES: QLQ-C30 AND FLIC AS EXAMPLES**

Madeleine T. King, Centre for Health Economics Research & Evaluation, University of Sydney, Sydney, NSW, Australia

This paper applies multilevel models for multivariate longitudinal data to the problem of equivalence of QOL scales. The proposed method assesses two aspects of equivalence simultaneously: 1) distinguishing between patients; 2) registering changes in QOL over time. Two cancer QOL instruments, QLQ-C30 and FLIC, illustrate the method. Five dimension scales were constructed from FLIC to measure global QOL, role function, emotional function, nausea and pain. The aim was to test whether these five scales gave equivalent results to QLQ-C30 scales with analogous content. The sample comprised 1041 QOL assessments on 138 heterogeneous cancer patients using FLIC and QLQ-C30 concurrently. Consider a set of repeated measurements on a single QOL scale, with an overall mean and a mean for each subject. Variations in the data can be decomposed into deviations of subject means from the overall mean and deviations of scores at different times from their subjects' means. The subject-specific deviations represent differences in QOL between patients, due for example to differences in health and personality. The time-specific deviations represent fluctuations in QOL, both real and perceived. A multivariate variance component model estimates the correlations between 1) the subject-specific deviations, and 2) the time-specific deviations of two QOL scales. These correlations for each pair of analogous scales from FLIC and QLQ-C30 are shown in the table. These suggest FLIC and QLQ-C30 distinguish between subjects similarly in each dimension. There is less equivalence in the way FLIC and QLQ-C30 register changes in QOL over time, although the correlation is moderate for each dimension.

Correlations	Global QOL	Role	Emotional	Pain	Nausea
Subject-specific	0.93	0.87	0.78	0.85	0.84
Time-specific	0.62	0.51	0.48	0.48	0.62

Abstract 1222

RESPONSIVENESS OF THE NATIONAL EYE INSTITUTE VISUAL FUNCTION QUESTIONNAIRE TO CHANGES IN VISION

Päivi H. Miskala, Center for the Submacular Surgery Trials (SST) Patient Centered Outcomes Subcommittee and SST Pilot Study Investigators, The Johns Hopkins University, Baltimore, MD

The National Eye Institute Visual Function Questionnaire (NEI-VFQ) measures vision-related quality of life in patients with vision impairment. Although the NEI-VFQ appears sensitive to between-patient vision differences, responsiveness to within-patient changes over time is unknown. This study investigated responsiveness of the NEI-VFQ subscales to changes in vision in patients with subfoveal choroidal neovascularization secondary to selected eye conditions, primarily age-related macular degeneration, enrolled in pilot Submacular Surgery Trials (SST). Data were combined from 3 pilot trials and correlation and regression methods were used to analyze changes in the NEI-VFQ scores and visual acuity from the 12-month to the 24-month visit using SAS. Eighty-seven of 327 patients enrolled in the pilot trials had an interview and visual acuities at 12 and 24 months. Median change in the overall NEI-VFQ score was a loss of 1.4 points (range -28.0,20.3). Median better eye visual acuity change was 0.0 lines (range -9.2,3.4). Eight of the 12 NEI-VFQ scales appeared responsive to changes in better eye visual acuity based on linear regression analysis ($p < 0.05$); general vision, near activities, distance activities, mental health, role difficulties, dependency, driving and peripheral vision scales. General health, ocular pain, social functioning and color vision scales were not sensitive to changes in better eye visual acuity. Similar results were obtained using weighted visual acuity of the 2 eyes with the exception of peripheral vision scale. Change in the worse eye visual acuity was not associated with changes in NEI-VFQ scales. Since several NEI-VFQ scales appear to be responsive to both differences and changes in visual acuity, it may be possible to extrapolate visual acuity and change in visual acuity when patients can be interviewed but visual acuity is not available. This analysis may help clinicians better understand which aspects of patients' lives are affected most by loss of central vision.

Abstract 1231

USING THE "SMALLEST REAL DIFFERENCE" AS A MEASURE OF SENSITIVITY TO CHANGE IN HEALTH-RELATED QUALITY OF LIFE SCALES: ADVANTAGES, PROBLEMS, AND SUGGESTED SOLUTIONS.

Peter Schuck, FBK Research Institute, Bad Elster, Germany

Recently it was suggested to use the so-called "smallest real difference" (SRD) as a measure of sensitivity to change/responsiveness [1]. The SRD is a confidence limit of the standard error of measurement (SEM) of the difference scores. The absolute values of the SEM/SRD per se are an indicator of measurement error (and as such complementary to the concept of reliability) and not a measure of sensitivity to change. In head-to-head studies with e.g. an effective intervention to induce change however, the percentage of patients, reaching the respective SRD criterion, could be used to compare the responsiveness of competitive instruments. In contrast to other sensitivity to change measures (e.g. "the" effect size) such an index would take the different reliabilities of the competitive instruments explicitly into account. The SRD concept is equivalent to the "reliable change index" (RCI), developed to assess "clinical significance" of outcome data for *single* patients in psychotherapy [2]. However, a patient here must not only have a pre-post difference score which is statistically different from measurement error (the RCI/SRD-criterion, see above), s/he must have returned with the post-treatment score to values of "normal functioning" of healthy people. Three different thresholds for the latter are discussed. Preferences for one or the other depends e.g. on the amount of overlapping of the two distributions. This will be presented, with a special focus on the possibility to use this approach as a substitute of an external criterion for change, which is often difficult to find. **References:** [1] Pfennings LE, van der Ploeg HM, Cohen L, Polman CH. A comparison of responsiveness indices in multiple sclerosis patients. *Qual Life Res* 1999;8:481-9. [2] Jacobson NS, Roberts LJ, Berns SB, McGlinchey JB. Methods for defining and determining the clinical significance of treatment effects: description, application, and alternatives. *J Clin Consult Psychol* 1999;67:300-7.

Abstract 1509

ABILITY OF THE CHILDREN HEALTH QUESTIONNAIRE (CHQ-CF87) TO DETECT DIFFERENCES IN QUALITY OF LIFE (QOL) BETWEEN CHILDREN ACCORDING TO THE TYPE OF CANCER

C RODARY, J LANDGRAF, C KALIFA, G LEVERGER, JC GENTET, BIOSTATISTICS AND EPIDEMIOLOGY, INSTITUT GUSTAVE ROUSSY, VILLEJUIF, FRANCE

OBJECTIVE : To answer the question : is the CHQ able to identify patients (pts) who have specific needs following completion of their treatment (TT), according to the type of cancer ? **INSTRUMENT :** the 2nd French version of the CHQ, the validation of which is ongoing, comprises 81 questions grouped into 8 dimensions : Physical Functioning (PF), Role/Social Limitations (RS), Self Esteem (SE), General Health Perception (GH), Bodily Pain (BP), General Behavior (BE), Family Activities (FA), Mental Health (MH). It provides an 8-score profile, ranging from 0 (very bad QoL) to 100 (very good QoL). **METHODS :** Children aged 9 to 19, suffering from 8 types of cancer were included. The CHQ was administered only once at the follow-up visit. Accrual of at least 300 children from 10 centers is expected. **RESULTS :** 261 pts have been included (May 2000). Compliance was high : 98%. The results concern 4 types of cancer with at least 30 pts per malignancy at the time of the analysis : Leukemia (LE),62, Non Hodgkin's Lymphoma (NHL),34, Wilms' tumor (WT),31, Bone tumors (BT),35. The median age is 13.5 years and the median interval since the end of TT is 4.5 years (range : 1.5 - 16). There was a significant difference between the 3 dimensions (Kruskall-Wallis test) : PF (respectively 90, 93, 97 and 81, $p < .001$), GH (respectively 64, 69, 72, and 59, $p < .01$) and SE (respectively 77, 81, 79, and 68, $p < .001$). For the type of cancer, the NB group had the highest scores and the BT group, the lowest. **DISCUSSION :** These results were expected but must be verified in a larger series of pts for the other 4 cancers. If this generic tool shows differences in QoL in pts, related to the cancer type, the next stage will be to evaluate the ability of CHQ to detect meaningful changes over time, in order to validate its role as a criterion of efficacy in oncology.

Abstract 1672

MEASURING RESPONSIVENESS IN QUALITY OF LIFE RESEARCH

Kevin W. Smith, New England Research Institutes, Watertown, MA

PURPOSE. Responsiveness, the degree to which an instrument measures change over time, is an important psychometric characteristic of quality of life (QOL) questionnaires. Responsiveness has traditionally been measured by an effect size. Effect sizes, however, have more to do with the efficacy of treatments than with the inherent properties of a QOL scale. The purpose of this research was to develop structural equation models (SEMs) to express responsiveness as the correlation between change in a QOL scale and change in a latent QOL variable. **METHODS.** Latent variable SEMs for the relationships among baseline, follow-up, and change scores were developed. The models were applied to 6-month change data for 181 patients with cardiovascular disease who completed the Multidimensional Index of Life Quality (MILQ) and for 89 HIV+ adults completing the Multidimensional Quality of Life Questionnaire for Persons with HIV/AIDS (MQOL-HIV). Retrospective assessments of perceived change and other indicators were used to identify the models. **RESULTS.** An SEM linking latent variables of baseline, follow-up, and change scores to observed scores is unidentified. Additional indicators or parameter constraints are needed to identify these models. Three approaches for identifying models were found. 1) Collect three indicators of change to form a just-identified model. 2) Treat responsiveness as the reliability of a change score and use the reliability formula to compute a responsiveness correlation. 3) Compute the correlation between change in a QOL scale and retrospective perceived change (rated from much worse to much better). Identified models will not always adequately fit the data. Because responsiveness correlations are a type of reliability coefficient, they depend on the variation in change occurring in a particular sample and will therefore vary from study to study. While difficult to estimate in some situations, responsiveness correlations are more accurate than effect sizes for assessing an instrument's ability to measure change.

RESPONSIVENESS OF HEALTH-RELATED QUALITY OF LIFE (HRQL) TOOLS AND CLINICALLY RELEVANT CHANGE IN PATIENTS WITH MUSCULOSKELETAL DISEASE

Barbara L. Spady, Maria E. Suarez-Almazor, Department of Educational Psychology, University of Alberta, Edmonton, AB, Canada

We compared the responsiveness and gains in quality-adjusted life years (QALYs) of different preference-based HRQL indices in patients with musculoskeletal disease. One hundred and seventy seven consecutive patients visiting a rheumatology clinic were invited to participate in the study. Self-response questionnaires were administered at baseline, 3, 6, and 12 months (m) with the following HRQL tools: the EuroQol (EQ-5D), the Health Utilities Index (HUI), and the SF-36. Preference-based weights were used to calculate the indices which give a utility value for health states (0=death and 1= perfect health): the EQ-5D weights are based on time trade-off methods, the HUI on standard gamble (SG), and the SF-36 has weights derived from visual analogue scales (VAS) and from SG. Patients were asked to rate their overall health compared to previous time periods and responses were categorized into 3 levels: better, same, and worse. One hundred and twenty patients completed the questionnaires at baseline and 6m, and 109 at baseline, 6 and 12m. Using ANOVA, there were significant differences in mean change scores and in QALYs gained for the EQ-5D, HUI, and SF-36 index scores for self-reported change in HRQL at 6m and 12m. For patients consistently reporting better HRQL over 1 year, QALYs gained were lowest for the SF-36 index based on standard gamble derived weights (0.04) and highest for the EQ-5D index (0.14). These findings have implications for the interpretation of clinically relevant change for different preference-based HRQL tools in patients with musculoskeletal disease.

Abstract 1545

THE ROLE OF PESSIMISM/OPTIMISM IN HRQOL AND MENTAL HEALTH IN CHRONIC HEPATITIS C (CHC) PATIENTS

Cheryl A. Moyer, Khozema Hussain, Robert Fontana, Steve Schwartz, Anna Lok, CHOICES, Department of Internal Medicine, University of Michigan, Ann Arbor, MI

The aim of our study was to determine the relationship between optimism/pessimism and both mental health and HRQOL in CHC patients. 220 CHC patients presenting to an academic medical center completed a survey assessing CHC risk factors, substance abuse history, HRQOL (SF-36), mental health (Brief Symptom Index), and expectations about CHC. Patients' expectations ranged from "I will continue to stay healthy" to "I will become sick very rapidly and I might die from liver disease in a few years." Responses were recoded into optimists (13.8%), realists (79.6%), and pessimists (6.5%). These categories were compared against demographics, disease severity, presence and number of active comorbid medical conditions, substance abuse history, mental health, and QOL using chi square, Kruskal-Wallis, and regression. Despite the small number of pessimists, pessimism was significantly associated with lower HRQOL scores on all SF-36 subscales except bodily pain ($p < .001-.03$). Pessimism was also significantly associated with elevated scores on the BSI subscales of somatization ($p=.023$), interpersonal sensitivity ($p=.016$), depression ($p=.006$), phobic anxiety ($p=.008$), paranoid ideation ($p=.003$), and the Global Severity Index ($p=.021$). Pessimism/optimism was not significantly associated with demographic factors, liver disease parameters (ALT levels, presence of cirrhosis) or substance abuse history. The number and presence of comorbid medical and psychiatric conditions were not associated with pessimism/optimism in univariate analysis. Self-reported health status on a scale of 1 to 5 (fair to excellent) was associated with pessimism/optimism at $p=.014$. We conclude that pessimism/optimism plays an important role in CHC patients' mental health, perceived QOL, and overall health status, regardless of their disease severity, comorbidities, or demographic characteristics. Further research is needed to assess its impact on disease progression.

Abstract 1546

DISTRESS AND IMMUNE FUNCTION IN CANCER PATIENTS RECEIVING A PHASE I VACCINE TREATMENT

Lorenzo Cohen, Cherylyn Savary, Carl de Moor, Stephen P. Tomasovic, Robert J. Amato, Department of Behavioral Science, U.T. M. D. Anderson Cancer Center, Houston, TX

Stress has been found to decrease immune function in cancer patients and social support may moderate this association. However, no previous research has examined the effects of stress and social support on patients' immune function against their own tumor (autologous killing). Stress, social support, and immune indices were measured in 30 newly diagnosed stage IV renal cell carcinoma patients prior to immunotherapy using an autologous tumor vaccine. Patients completed the Perceived Stress Scale and the Interpersonal Support Evaluation List. We measured cytotoxicity to K562 target and autologous tumor target, and CD4, CD8, and NK cell numbers. Multiple regression analyses revealed that, after controlling for disease status variables (number of metastases, site of metastases, and calcium levels), perceived stress was negatively associated with cytotoxicity to K562 cells ($p < 0.003$) and to autologous tumor cells ($p < 0.005$), and positively associated with CD8 cell numbers ($p < 0.008$). Importantly, perceived stress was still negatively associated with cytotoxicity to autologous tumor cells even after controlling for cytotoxicity to K562 target cells and the other covariates ($p < 0.02$). This suggests that perceived stress was associated with cytotoxicity to autologous tumor cells independent of a more general immune response. Social support was also a significant moderator of the perceived stress/cytotoxicity to autologous tumor relationship ($p < 0.05$), in that as the perceived stress levels increased, autologous cell killing decreased, but less so in patients reporting high levels of social support. These results may be especially relevant to the clinical course of disease progression given that these patients will receive an autologous tumor vaccine to boost their immune systems ability to fight their own tumor. Importantly, psychosocial interventions may affect the clinical course of disease in cancer patients receiving vaccine treatment.

Abstract 1765

EFFECT OF SENSE OF COHERENCE ON ANGIOPLASTY PATIENT QUALITY OF LIFE BOTH PRE AND POST-SURGICALLY

Colleen M. Renier, Carl E. Heltnie, Jeanette A. Palcher, Robert L. Tilden, Division of Education and Research, St. Mary's/Duluth Clinic Health System, Duluth

Background: A great deal has appeared in recent literature regarding the relationship between Sense of Coherence (SOC) and Quality of Life (QoL). This study was designed to aid in the interpretation of longitudinal QoL, by assessing the effect of SOC. Methods: To better understand the effect of SOC on QoL for patients receiving Percutaneous Transluminal Coronary Angioplasty (PTCA), the Health Status Questionnaire (HSQ) 2.0, a slightly modified version of the SF-36, was given to cardiology patients scheduled to have their first PTCA. The same form was then mailed to the patients at six, twelve, and 36 months post-procedure. 302 PTCA patients entered the study, with 82.1% responding at six months, 78.8% at twelve months, and 64.6% at 36 months. 204 of the patients were male, and 98 were female. Comparisons were made by calculating z-scores (the number of standard deviations from age/gender specific norms). Results: Significant positive correlations existed between SOC and all eight of the HSQ 2.0 domains of self-reported health status, at all four time periods ($p < .05$ for General health at 36 months, and $p < .01$ for all other comparisons). Repeated measures analysis of variance found significant relationships between SOC and all eight of the domains ($p < .01$), while the interrelationship of time and SOC was only significant for Emotional well-being ($p < .01$). Additional analysis of change from baseline found no significant relationships between SOC and any of the eight domains, at any of the three followup time periods. Conclusions: While there is undoubtedly a strong relationship between SOC and QoL, the effect is static, effecting HSQ 2.0 scores consistently across time. Therefore, while SOC is an important factor when assessing how well a patient is doing, at a given point in time, compared to other patients of the same age and gender, it has little effect when comparing patients' levels of improvement over time.

Abstract 1732

DETERMINANTS OF THE QUALITY OF LIFE OF DIABETES PATIENTS TREATED IN DIFFERENT MEDICAL INSTITUTIONS

Matthias Rose, Herbert Fliege, Department of Psychosomatic Medicine, Charité, Berlin, Germany

The HRQL is, in addition to blood sugar control, one of the most important outcome variables in the treatment of diabetes. Using a large random sample of diabetes out-patients, we examined which factors determine their QOL. — A total of 625 patients from various randomly chosen treatment facilities were questioned (187 pat. from 28 GP's, 264 pat. from 3 diabetes specialists, 174 pat. from a university clinic), 224 type I and 401 type II diabetic patients. 10 standardized German Questionnaire were used, including the WHOQOL-Bref, a coping inventory and a Knowledge-Test. In addition, data was gathered from the physicians. — On average, the HRQL of diabetes patients is reduced in all dimensions measured in the WHOQOL-Bref when compared to a representative sample of healthy individuals. Marked differences were found, however, in the various sub-groups: for example, patients without secondary illnesses hardly experienced any reduction in their HRQL. A regression analysis revealed (R^2 corr. = .28, $p < 0.001$) that a patient's reporting of physical well-being is dependent not only on the presence of secondary illnesses ($b = -.37$, $p < .001$), but, also on the presence of an active coping behaviour ($b = .20$, $p = .001$) and on the level of specialization of the treatment facility ($b = .19$, $p < .05$). In the other HRQL dimensions, psychological variables played an even more important role in the HRQL. Age, gender, length of illness, illness-specific knowledge, or treatment attitude of the physician were of little significance. — If the QOL of patients with chronic illnesses is to be a goal in medical treatment, it is important to become acquainted with those factors which influence it: In the case of the diabetes patients, their state of well-being was influenced as much by their illness coping strategies as by the degree of specialization of their treating physician. Patients felt better in those facilities that encouraged more active forms of dealing with their illness and in which intensified forms of therapy were more often used.

Abstract 1193

THE IMPACT OF ETHNICITY AND TREATMENT ON HEALTH-RELATED QUALITY OF LIFE (HRQOL) OF AFRICAN AMERICAN AND CAUCASIAN MEN WITH PROSTATE CANCER: DATA FROM CAPSURE

Deborah P. Lubeck, Gary D. Grossfeld, Scott C. Flanders, Peter R. Carroll, Department of Urology, University of California, San Francisco, San Francisco, CA

Introduction and Objectives: To compare clinical characteristics, treatment choices, sociodemographic features and baseline and follow-up HRQOL of African American (AA) and Caucasian men with prostate cancer. **Methods:** 1178 newly diagnosed patients (82% Caucasian, 12% AA) from the CaPSURE database (a national, longitudinal registry of patients with prostate cancer) were studied. Baseline cancer stage, grade, serum PSA, comorbidity, treatment selection and baseline and post-treatment HRQOL were assessed using the SF-36 and the UCLA Prostate Cancer Index. The impact of ethnicity on these outcomes was determined. **Results:** Significant differences ($p < 0.001$) were noted between ethnic groups. AA men were younger, less educated and had lower income than Caucasians. AA men were more likely to have high-stage (i.e. N+ or M+) (10% vs. 1%) disease and higher pretreatment serum PSA (12.7 ng/ml vs. 7.5 ng/ml). AA men were more commonly treated with androgen deprivation (37% vs. 26%). Clinically and statistically significant differences in baseline HRQOL were noted in AA men who had poorer: General Health, Physical Function, Bodily Pain, Role Emotional Function, Sexual Bother, Self Esteem, and Health Distress. While both groups improve in HRQOL after treatment, AA had worse outcome in several domains controlling for treatment: Bodily Pain, Self Esteem, Health Distress, Physical Function, General Health, Role Emotional, Urine Function and Bother, Bowel Function and Bother, and Cancer Interference. These differences were statistically and clinically significant. **Conclusions:** Significant differences exist in baseline clinical presentation and HRQOL between AA and Caucasian men with prostate cancer. Differences remain after treatment. The reasons for these differences require further study. Interventions, which might reduce these differences, need to be developed.

Abstract 1677

ARE HEALTH STATUS AND QOL ASSOCIATED WITH COMPLEMENTARY THERAPY USE IN TRANSPLANT PATIENTS?

Barbara J. Matthees, Cynthia R. Gross, Mary J. Kreitzer, Kay Savik, Puree Anantachoti, Marshall I. Hertz, Academic Health Center, Minnesota State University/ Moorhead, Moorhead, MN

Patients post lung transplant (Tx) experience complex physical psychological and emotional side-effects. They may choose to alleviate these chronic problems through the use of various complementary and alternative therapies (CAM). This descriptive study was conducted to estimate the prevalence of CAM use and determine if CAM users differ from non-users with respect to health status, QOL, treatment satisfaction or adherence. A mailed self-report survey was sent to 145 lung Tx recipients 6 months or more post-transplant. 99 patients responded. Respondents were: female 58%; mean age of 51 yrs; time from transplant median = 50 mos, range 9-123; with heart-lung (9%), single lung (55%), or bilateral single lung (36%) transplants. Patients were asked about use of 23 different CAM therapies and 88% reported using at least one. 71% reported using more than one CAM therapy (median = 2). Only 44% of those using CAM reported discussing CAM with their providers. Prayer (69%), support group (43%), and nutritional supplements (21%) were the most common. Overall, few differences were found between CAM users and non-users. CAM users, in general, were more likely to be college educated. Women were more likely to report use of prayer. Support group participants reported having higher steroid-related side effects. Other CAMs users reported more Illness-Intrusion, more symptoms and depression on the CES-D scale. Patients with high symptom burden are likely to use particular forms of CAMs. Physicians may assume that most Lung Tx patients are using some CAM therapies and need to explore potentials for interaction or enhancement with standard therapy.

Abstract 1613

ALWAYS A NORMAL DISTRIBUTION OF HRQOL? ASSUMPTIONS ABOUT THE DISTRIBUTION OF THE LATENT VARIABLE INFLUENCE IRT RESULTS

Jakob B. Bjorner, Department of Psychology and Sociology, National Institute of Occupational Health, Copenhagen, Denmark

IRT analyses of data on HRQOL have used either Rasch models or general IRT models (Generalized Partial Credit or Graded Response). Whereas Rasch type models do not rely on distributional assumptions, other IRT models typically assume that the latent HRQOL variable has a standard normal distribution. Many HRQOL scales are skewed, which may reflect both ceiling problems and skewness in the latent distribution. We performed a simulation study concerning the potential impact of such skewness on the results of IRT analyses. We simulated data using a general polytomous IRT model and typical values of item parameters seen in HRQOL studies. We simulated various degrees of skewness in the latent distribution of HRQOL. We then estimated the item parameters and the latent HRQOL scores using standard methods and assuming normality. Finally, we checked whether the estimates recovered the correct simulated values. We found that in case of skewness, the item slope estimates were typically biased downwards, while the threshold estimates were biased towards the mean of the thresholds for that item. However, considerable bias in the other direction was also seen. The average difference between simulated and estimated values of the latent HRQOL scores was larger than expected from the model. Estimates were biased downwards for simulated HRQOL below the mean (0), and biased upwards otherwise. Analyses of skewness and kurtosis of the estimated HRQOL scores considerably underestimated the skewness and kurtosis of the simulated scores. We conclude that HRQOL researchers using general IRT models need to pay attention to their assumptions about the distribution of HRQOL. Wrong assumptions will bias the item parameter estimates and the IRT based HRQOL scores. A post hoc analysis of skewness and kurtosis of the estimated HRQOL scores underestimates the problems. We present two other approaches to checking the assumptions about the latent distribution.

Abstract 1715**UNBIASED QUALITY OF LIFE MEASUREMENT ACROSS LITERACY LEVELS AND ADMINISTRATION MODES IN SPANISH-SPEAKING CANCER/HIV PATIENTS**

Elizabeth A. Hahn, David Cella, Center on Outcomes, Research and Education (CORE), Evanston Northwestern Healthcare, Evanston, IL

In order to accommodate the needs of patients with low literacy, and to provide maximum flexibility in administration of quality of life (QOL) instruments, alternative modes of administration are desirable. The benefits of such flexibility, however, must not be at the expense of obtaining unbiased measurement of QOL. A multicenter study of cancer and HIV patients included an aim to test the psychometric properties and statistical equivalence of the Spanish language version of the Functional Assessment of Cancer Therapy (FACT) across literacy levels (low versus high) and administration modes (self versus interview). High literacy patients were randomly assigned to self ($n=214$) or interviewer ($n=248$) administration of the FACT, while interviews were conducted with all low literacy patients ($n=198$). An extension of the Rasch measurement model for rating scale data was used to evaluate the possibility of differential item functioning. Separate item difficulty calibrations were obtained for the three patient groups for each of the FACT subscales (physical, social, emotional and functional well-being), and these were evaluated for invariance using a 95% confidence interval approach. The results of three pairwise comparisons for each subscale suggest that common item difficulty calibrations are appropriate for most FACT items. In other words, we did not detect evidence of systematic measurement bias across literacy levels or administration modes. Literacy-fair QOL evaluation for patients with cancer or HIV, and technical equivalence of modes of administration, will permit unbiased assessment of the impact of cancer/HIV and their treatments on Spanish-speaking patients from diverse backgrounds.

Abstract 1701**THE CONTRIBUTION OF RESPONSE PATTERNS TO THE RESPONSIVENESS OF MULTI-ITEM SCALES: A COMPARISON OF ITEM RESPONSE THEORY (IRT) VERSUS SUMMATIVE SCORING FOR MEASURING CHANGE USING THE SF-36 PHYSICAL FUNCTION SCALE.**

Kitty S. Chan, Department of Health Policy and Management, Johns Hopkins School of Hygiene and Public Health, Baltimore, MD

As multi-item scales are increasingly relied upon to evaluate changes in health, it is important that scale scores accurately reflect underlying change rather than instrument characteristics. This study compared summative versus IRT (Graded Response) scoring methods in reflecting one-year change in physical function using the SF-36 Physical Function (PF) Scale. Subjects were participants in an asthma clinical trial ($N=820$, baseline 40-90% FEV1 predicted). The influence of scoring method on size of score change, score variance, and correlation with pulmonary function were examined. Results indicate IRT change scores are less likely to be affected by item difficulty distributions and more closely related to underlying pulmonary changes than summed change scores. The correlation of changes in FEV1 with change in PF scores is stronger for IRT scores ($r = 0.21$) than for summative scores ($r = 0.13$). Meanwhile, the size of score changes was more closely related to baseline PF (summed) scores for summative change scores ($r = -0.40$) than for IRT change scores ($r = -0.20$). The clustering of item difficulty in the moderate range of physical function likely contributed to the inflated gains for low baseline scores and limited gains for high baseline scores in the summative method. These results suggest that scores based upon item response models offer a better approach to evaluating underlying change than summative scores, especially for instruments with uneven item difficulty distributions.

Abstract 1578**EVALUATION OF THE WHOQOL-BREF IN SIX COUNTRIES: A CROSS-CULTURAL EVALUATION USING RASCH ITEM RESPONSE THEORY (IRT) ANALYSIS**

Donald M. Bushnell, Marianne Amir, Department of Analytic Services, Health Research Associates, Inc., Seattle, WA

The WHOQOL-Bref is a generic quality of life measure, developed simultaneously in many cultures and languages. It consists of the following 4 independent subscales: Physical, Psychological, Social, and Environment. Standard psychometric procedures have been employed for validating this instrument in each country. The Longitudinal Investigation of Depression Outcomes (LIDO) is a currently ongoing observational study taking place in Israel, Spain, Australia, Brazil, the United States, and the Russian Federation. With data from this study, it is our goal to further compare the WHOQOL-Bref across these six countries aiming at evaluating the extent to which the item characteristics remain stable at each site and examining whether the multiple aspects of the WHOQOL-Bref form hierarchical and unidimensional indices that are reproducible across cultures. Data gathered from a sample of patients exhibiting depressive symptoms, but not necessarily diagnosed with major depression, were used. A series of Rasch analyses were performed to address the hierarchical order and spacing of items based on item calibration and standard error estimates and the unidimensionality of the scales based on a goodness-of-fit analysis. The combined results of the Rasch-IRT analyses support the WHOQOL-Bref subscales as hierarchical indices and reproducible across these six countries. The average item-calibration estimates incorporate good ranges of logit values with good item spacing (Physical: 1.0 to -1.5; Psychological: 0.8 to -1.1; Social: 0.8 to -0.7; Environment: 1.2 to -1.0) similar across the countries. The goodness-of-fit analyses indicate acceptable unidimensionality within each subscale. No standardized fit statistic values fell above 2.0 (indicating items not closely related to the overall construct) or below -2.0 (indicating items with redundant content).

Abstract 1148**THE AMSTERDAM LINEAR DISABILITY SCORE PROJECT (ALDS): CALIBRATION OF AN ADL-ITEMBANK TO MEASURE FUNCTIONAL STATUS IN CLINICAL RESEARCH.**

Robert Lindeboom, Rebecca Holman, Rob J. DeHaan, Rien Vermeulen, Department of Clinical Epidemiology and Biostatistics, Academic Medical Center, Amsterdam, The Netherlands

In clinical studies, the effect of a therapy on the functional status or QoL of patients is frequently assessed using summative rating scales, which have an ordinal, rather than an interval level of measurement. Moreover, these scales have a limited range of measurement (e.g. basic activities of daily living or instrumental ADL or some of both) and show poor precision in more able or disabled patients (floor and ceiling effects). The goal of the ALDS-project is to calibrate a large ADL-item bank using IRT statistical models. The item bank consists of 190 items and can be used to assemble disease specific and/or ability tailored functional scales. The scales that are assembled from the item bank share a common metric. Therefore, it is possible to compare the functional status of patient groups with different levels of ability and possibly different diagnoses, as well as the effectiveness of therapies from various medical specialties. Subsets of the item bank (e.g. 30 items) are administered to large groups of patients (> 200) with a wide variety of chronic diseases. Using anchored item designs routinely applied in educational measurement, all items are placed on the same linear (interval level) ruler of ADL-functioning without burdening patients with the total set of items. We will present the calibration results of the first 75 items of the item bank. The calibration of the resulting 115 items is ongoing. Due to misfit, differential item functioning and local dependence, we expect that about 110 to 120 calibrated items will remain in the item bank after the study is completed. The generic item bank can be used in clinical trials and prognostic studies and cost-effectiveness studies.

Abstract 1418

QUALITY OF LIFE IN CORONARY PATIENTS WITH PREOPERATIVE HEART FAILURE BEFORE AND AFTER MYOCARDIAL REVASCULARISATION

Bogoljub B. Mihajlovic, Svetozar R. Nicin, Djordje Jakovljevic, Ninoslav D. Radovanovic, Department of Cardiovascular Surgery, Institute of Cardiovascular Diseases, Sremska Kamenica, Serbia, Yugoslavia

Background: Conventional surgery in coronary patients with preoperative heart failure, ejection fraction (EF) lower than 30% is an alternative to heart transplantation. Therefore, the good postoperative results, as well as the quality of life (QOL) after myocardial revascularisation are of utmost importance. The aim of this study was to compare QOL between patients with preoperative EF lower or equal to 30% and to those with EF > 30% before and 6 months after myocardial revascularisation. Methods: We analysed 360 consecutive patients, 300 (83%) with EF > 30% and 60 (17%) with EF lower or equal to 30%. For QOL assessment we used a self-designed questionnaire consisting of four domains: physical status, mental status, social interaction and self-perception of health. For statistical analysis, before and six months after operation, we used pairwise and independent samples t-test of integral numerical value QOLI-NS, that represents integral overall numerical value of QOL. Results: QOL improvement after myocardial revascularisation was statistically significant in both groups of patients (p<0.001). Six months after surgery there was no difference in QOL between patients with preoperative heart failure (EF<=30%) and patients with EF>30%. Conclusion: Our results confirm benefit from myocardial revascularisation in coronary patients with preoperative heart failure.

QOLI-NS	Results		EF>30 vs. EF<=30 (p)
	EF<=30	EF>30	
Before op.	41.6±21.2	39.8±16.6	n.s.
6 m. after op.	82.6±15.8	84.3±14.0	n.s.
Before : after (p)	<0.001	<0.001	
QOL improvement	41.1±24	44.5±19.8	n.s.

Abstract 1652

PRE-PROCEDURE ANGINA FREQUENCY PREDICTS COST-EFFECTIVENESS OF CORONARY REVASCULARIZATION

John A. Spertus, Brent D. Bliven, Philip G. Jones, Joseph F. Tooley, Kasem S. Akhras, James O'Keefe, Daniel Lips, A M. Borkon, Department of Cardiology, Mid America Heart Institute of Saint Luke's Hospital, Kansas City, MO

Cost divided by quality-adjusted life-years (QALYs = survival X utility) determines the cost-effectiveness of coronary revascularization. Utilities range from 0-1 where 0 represents death & 1 equals the best possible health state. The greater the improvement in utilities, the more cost-effective the procedure, assuming equivalent costs of the procedure in patients with differing frequencies of angina. This study investigated the relationship between angina frequency, assessed by the Seattle Angina Questionnaire, & the benefits of revascularization as measured by the Health Utilities Index (HUI). This was a prospective, observational cohort study of 317 consecutive patients. Questionnaires were administered at baseline & monthly for 6 months. Subjects were divided into 3 clinically meaningful ranges (daily angina, weekly angina, monthly/no angina). The change in HUI scores from baseline to 6 months reflects the benefit in QALYs from revascularization. Patients' utilities at baseline were inversely related to the frequency of angina. The greatest benefits from revascularization were achieved in patients with the most frequent angina prior to their procedure. Analysis of variance revealed that these differences were statistically significant (p<.001). Patients with more frequent angina realize greater utility benefits from coronary revascularization. These differences in utility suggest that it may be more cost-effective to treat patients with more frequent angina.

ANGINA FREQUENCY	HUI BEFORE REVASCULARIZATION	CHANGE IN UTILITY
Daily (n=61)	.54	.13
Weekly (n=100)	.68	.06
Monthly/none (n=156)	.78	.006

Abstract 1702

INTERNATIONAL COMPARISON OF BASELINE AND IMPROVEMENTS IN HEALTH STATUS FROM PERCUTANEOUS REVASCULARIZATION

Kasem S. Akhras, Michael Coen, Joseph F. Tooley, Philip G. Jones, Brent D. Bliven, John A. Spertus, Searle, Skokie, IL

Different practice patterns may alter the types of patients referred for revascularization and the benefits they receive. We examined differences in symptoms, functioning and quality of life in 5 countries participating in the EXCITE Trial (EXCITE=Evaluation of Oral Xemilofiban in Controlling Thrombotic Events). The Seattle Angina Questionnaire (SAQ, range 0-100 with higher scores = better) was administered to 683 patients at baseline and 6 months. The baseline and change scores were examined across Germany (n=120), France (n=100), England (n=135), Canada (n=106) and the USA (n=222). ANOVA and general linear models examined whether country of origin was associated with differences in SAQ scores. Baseline SAQ angina frequency scores ranged from 43 in England to 61 in France (p<0.0001) and changes ranged from 19 in France to 30 in Canada (p = .04). Baseline SAQ physical function scores ranged from 54 in England to 70 in the US (p<0.0001) and changes varied from 9 in the US to 22 in England (p=0.0002). Baseline SAQ quality of life ranged from 40 in England to 53 in the US (p<0.0001) and changes varied from 15 in France to 31 in Canada (p<0.0001). Baseline SAQ treatment satisfaction ranged from 83 in France to 89 in the US (p<0.0001) and changes varied from 1 in the US to 5 in Canada (p=0.001). Once baseline demographic factors, clinical variables and baseline health status were accounted for, no differences between countries were seen for relief from angina but statistically significant differences in changes in score were seen for physical limitation, treatment satisfaction and quality of life. Although patients' symptoms, functioning and quality of life differ at the time of referral for percutaneous revascularization among different countries, less variation in the benefits (changes in SAQ scores) of revascularization were observed. Further investigation is needed to verify these findings and to examine the factors responsible for the observed differences in outcome.

Abstract 1642

SF-36 AS A SCREENING INSTRUMENT TO DETECT SYMPTOMS AMONG HYPERTENSIVE PATIENTS

Steven R. Erickson, Brent C. Williams, Larry Gruppen, University of Michigan, Ann Arbor, MI

Purpose: Determine if the SF-36 can detect disease-specific symptoms among treated hypertensive patients. Subject sample and methods: Cross-sectional study used surveys mailed to patients' homes. Consecutive hypertensive patients over age 30 years attending either a general medicine or hypertension clinic during three months (n=222). Patients had documented primary hypertension, prescribed antihypertensive medications, and no other symptomatic conditions or therapies. Measures included the SF-36, Symptom Distress Checklist (SDC, list of 51 HTN specific symptoms, frequency, and distress, reduced to 14 symptom factors). Patient and treatment characteristics included blood pressure, length of diagnosis, age, gender, race, income, education, length of current therapy. Data were analyzed using logistic regression. The dependent variable was the presence or absence of any symptom in each of the 14 SDC factors prevalent among hypertensives. The independent variables were the SF-36 domains, controlling for patient and treatment characteristics. Logistic regression was used to determine which of the SF-36 domains were most consistently and strongly associated with symptoms, examining the area-under the ROC curve, the beta-coefficients, and the p value. Summary of Results: Among SF-36 domain scores, the General Health (GH) and the Physical Functioning Summary (PFS) were best overall predictors of symptoms. At predicted probability cutoff levels of 10 to 40%, adequate to detect most symptomatic patients (increased sensitivity), 82 to 98% of patients would screen "in" for the presence of a disease-specific symptom. As the predicted probability cutoff level is lowered (increased specificity), fewer symptomatic patients would be identified by the GH and PFS, but fewer patients would screen "in." GH and PFS were useful in detecting symptoms common in treated hypertensive patients but the sensitivity and specificity is somewhat lower than desired for clinical use.

Abstract 1624**TREATMENT OF CHRONIC HEART FAILURE (CHF) : IMPROVING SURVIVAL BUT NOT HEALTH-RELATED QUALITY OF LIFE (HRQL) ?**

Olivier Chassany, Isabelle Mahe, Martin Duracinsky, Unite de Recherches Therapeutiques, Hopital Lariboisiere, Paris, France

Treatments such as ACE inhibitors and beta-blockers reduce mortality in CHF. A parallel demonstration of a HRQL improvement would appear of good added-value. A critical appraisal of published controlled trials was performed using a checklist. Although a reduction of all cause mortality and NYHA score was demonstrated in a few large studies, none was able to show a clear and unbiased HRQL improvement with active treatment vs. placebo. Several issues that would help to have confidence in the results are not presented or missing: hypotheses of changes in HRQL scores and power estimation; justification of the choice of questionnaires (e.g. 6 questionnaires were used in a trial and presented as HRQL, some of them being a dyspnoea score or a global evaluation of change, with no explanation whether they were covering different concepts, and without selecting the most important questionnaire for the primary analysis); validation data (e.g. a trial used 14 subscales taken from 3 different validated questionnaires. Is the resultant questionnaire still valid ?); statistical analysis plan; description and imputation of missing data; statement that the analysis is in intent to treat. In the majority of trials, small differences were observed for a few subscales among many, or only at some assessments over time. When the difference reaches statistical significance, it is only because of the large sample, but it has no clinical signification, e.g. social function score (range: 1-4): 1.54 (active treatment) vs. 1.45 (placebo). Even for the physical dimension which may be of importance in CHF, small non significant effect sizes are observed. Some explanations proposed for these results are : non severe patients included, questionnaire not reliable or responsive. There is concern to conclude that despite attempts of publications and general reviews to enhance the modest results, there is no definite demonstration that HRQL is improved by treatment which reduces mortality in CHF.

Abstract 1619**DEVELOPMENT OF A CONCEPTUAL MODEL OF HEALTH-RELATED QUALITY OF LIFE (HRQL) FOR PEOPLE WITH DEMENTIA**

Sarah C. Smith, Joanna Murray, Beth Foley, Joanne Cook, Sube Banerjee, Donna L. Lamping, Health Services Research Unit, London School of Hygiene and Tropical Medicine, London, UK

The importance of a clear conceptual model in the development of a new outcome measure is well established, as comprehensive assessment of validity requires that the construct to be measured is clearly defined and described. In dementia the conceptual model is particularly important as the potential for measurement bias presents unique challenges to reliability and validity. Progressive decline of cognitive function in dementia may mean that the skills required to report on subjective phenomena such as HRQL are compromised or absent. It may be necessary to adopt a deliberately flexible approach about proxy or self-report and the method of administration of the instrument. Moreover, the nature of the condition may mean that the experience of HRQL in people with dementia is unique and is not necessarily captured by generic models of HRQL. Both top down (constructs obtained from relevant literature) and bottom up (new qualitative data) approaches were used in parallel to develop a new conceptual model for HRQL in dementia. A preliminary version of the model was derived from both the generic and disease specific HRQL literature and was additionally informed by expert opinion of clinical researchers working with people with dementia. This version consisted of 7 domains. On the basis of the content analysis of 10 qualitative interviews with people with dementia and their carers the preliminary model was refined to 5 domains. The refined model formed the basis of the second stage of qualitative interviews, which are currently in progress. Analysis of these interviews will inform the final refinement of the model. This final conceptual model will be used to develop a new instrument to measure HRQL for people with dementia, which will then be field tested for acceptability, reliability, validity and responsiveness.

Abstract 1228**HOW BAD ARE STATES OF DEMENTIA?**

Steven M. Albert, Gertrude H. Sergievsky Center, Columbia University, New York, NY

Withdrawn

Abstract 1301**IMPACT OF RIVASTIGMINE ON COSTS AND TIME SPENT CARING FOR ALZHEIMER'S DISEASE PATIENTS**

Deborah B. Marin, Edward H. Snyder, A. Brett Hauber, Linda S. Deal, Kenneth L. Davis, Department of Psychiatry, Mount Sinai Medical Center, New York, NY

Alzheimer's disease (AD) places a significant burden on healthcare systems world-wide. As new treatments are developed, their cost effectiveness is often assessed to help healthcare professionals make informed decisions. In addition to the more common practice of assessing direct medical costs, indirect costs including time spent caregiving should also be evaluated. In this study, the potential savings in indirect costs due to treating patients with the dual cholinesterase inhibitor rivastigmine Exelon® were examined. Results from two 26-week placebo-controlled trials have demonstrated the clinically relevant and statistically significant efficacy of rivastigmine (6-12mg/day) compared with placebo, for cognition, ADL and global functioning. By delaying disease progression from one stage of AD to the next, significant savings in indirect costs would be anticipated. Data collected in a prospective, observational study of AD patients and caregivers was used to establish the relationship between disease severity (according to MMSE score) and time spent caregiving (according to the 5-item CAS score). A significant correlation was observed between the two scores (n=43, r=-0.56, p<0.0001) demonstrating that as the disease progresses more time for supervision and help from caregivers is required. To estimate the effect of rivastigmine on economic burden of caregiving, we multiplied the difference in total caregiver hours between two disease stages by the number of days by which progression to a more severe stage of the disease is delayed due to treatment with rivastigmine. We then multiplied the hours saved by the average hourly wages for a nurse's aide (according to Rice et al 1993, adjusted to 1997 dollars). The reduction in time spent caregiving reached 691 hours over two years for caregivers of mildly demented patients (MMSE 21-30). In monetary terms, this time saved caregiving translated into a total saving of approximately \$8,808 over a two year period.

Abstract 1762**THE MEASUREMENT OF CAREGIVER QOL AND US ADAPTATION OF THE ALZHEIMER'S CAREGIVERS QUALITY OF LIFE INSTRUMENT (ACQLI).**

Meryl Brod, Doward C. Lynda, Stephen McKenna, Diane Whalley, The Lewin Group, San Francisco, CA

Caregiving for demented patients greatly impacts caregiver QoL. The ACQLI, based on a needs-based model, is a simple 30-item measure designed to assess QoL of these informal caregivers. It was developed simultaneously in 5 countries, was shown to have excellent psychometric properties and to be acceptable to respondents. This paper reports on the conceptual issues of measuring QoL for caregivers of demented patients and the adaptation and psychometric testing of the US-English version. The adaptation process involved three stages: translation, field-testing and psychometric testing. Translation was conducted by a panel of five US nationals who were of average or below average educational level. Face and content validity were tested by interviewing 15 caregivers. The final stage was a postal survey of 61 caregivers, assessed on 2 occasions, designed to establish test-retest reliability, internal consistency and construct validity (known groups and convergent). The measures were successfully translated with minor alterations made to render language or spelling suitable to US respondents. Twelve items required translation for conceptual equivalency. Field-testing indicated that the US-ACQLI is acceptable, relevant and easy to complete. It has excellent test-retest reliability (.91) and internal consistency (time 1: .90, time 2: .91). It was able to distinguish between caregivers reporting excellent, good or poor self-perceived general health ($p < .05$) and was found to relate as expected with the dimensions of the Nottingham Health Profile (moderate/high correlations and emotional reactions, social isolation and energy level, low correlations with pain and physical mobility). The ACQLI has been shown successfully adapted into US-English comparable to the original UK version. The US version is now suitable for use in clinical trials and community based studies to assess QoL and the impact of interventions on caregivers.

Abstract 1640**CLINICAL AND NON-CLINICAL DETERMINANTS OF CAREGIVERS' WELL-BEING (WB) IN ALZHEIMER DISEASE (AD)**

Benoit Arnould, Koen Torfs, Christine De La Loge, Christine De La Loge, Mapi Values, Lyon, France

Although caregivers' burden has been assessed in AD studies, little attention has been paid to their well-being (WB). This work evaluates caregivers' WB and studies its links with patients' clinical status, socio-demographics (SD), and medical resources use (MRU). The Psychological General Well-Being (PGWB) was proposed to the caregivers of 643 AD patients. It provides an index and a six-scales profile. Patients were assessed using a battery of scales, including ADAS-cog, DAD/ADL, MMSE, and a Dependence scale. SD and MRU were also collected. Correlation coefficient, ANOVA, and Multiple Factorial Analyses (MFA) were performed. The analyses include 591 caregivers (mean age 62, 58.9% women). The mean PGWB Index (PGWBI) was 71.2, which is in-between reference values in populations reporting - or not - co-morbidity (respectively 76.3 and 69.1). Significant links were found between PGWBI and both disability (DAD/ADL, $p < 0.0001$) and Dependence ($p = .007$) scales, but not with the cognitive scales (ADAS, MMSE). In addition, the need for round-the-clock supervision was correlated with the caregivers' depression score ($p = 0.007$). SD significantly linked to WB were: patient's gender ($p < 0.0001$), caregiver's age ($p < 0.0001$), patient and caregiver difference in age ($p < 0.0001$), time spent ($p = 0.004$), and caregiver living with or married to the patient ($p < 0.0001$). No link was found between caregivers' WB and MRU. MFA confirmed these results. The PGWB is well validated, and the large sample ensures sufficient power. The results show that caregiver's WB is linked to the patient's functional ability level, more than its cognitive performance. In addition, the relationship between the caregiver and the patient seems to have greater impact on caregiver's WB than the level of MRU. Analyses on longitudinal data will help to understand the way caregivers may benefit from an optimisation of treatment.

Abstract 1481**THE IMPACT OF PHYSIOTHERAPY IN LOW BACK PAIN PATIENTS**

Pedro L. Ferreira, Joao N. Gil, School of Economics, University of Coimbra, Coimbra, Portugal

The outcomes of physiotherapeutic care reveal the effect that physiotherapy has on the quality of the life of their patients. This research aimed at assessing the health-related quality of life of 91 patients suffering from low back pain and the role of physiotherapy in their recovery. A longitudinal study was performed in patients with diagnosed problems who went through ambulatory treatment in seven healthcare institutions, three public hospitals and four private hospitals. To carry out this study we used the Portuguese version of the MOS SF-36 as a generic measure and the Jette's Functional Status Questionnaire as a specific measure. The Beck Depression Index was also used. In the sample 73.8% of the patients were female, the majority (82.6%) were married and 66.3% had an education lower or equal than basics. The mean age was 45.3 year, from a minimum of 20 to a maximum of 78. We were able to evidence that non-specific low back pain problems affected patients in several aspects of the generic health status and functional health, and that patients' health improved four weeks and twelve weeks after the beginning of the treatment. This improvement was evident in more than pain and physical function dimensions; in fact, only social function, role function and social interaction were not significant when means were compared within the three-time period.

Abstract 1743**THE PATIENT-GENERATED INDEX IN RHEUMATIC DISEASES: CLINICAL CORRELATES AND RELATIONSHIP WITH GENERIC AND DISEASE-SPECIFIC QOL MEASURES. THE RHEUMAQOL STUDY.**

Marco Antivalle, Mariagrazia Marrazza, Salvatore Santandrea, Massimo Pagani, Rheumatology Unit, Azienda Ospedale L. Sacco - Polo Universitario, Milano, Italy

PURPOSE: to evaluate the clinical determinants of individual QOL in rheumatic diseases, and the relationships between different methods of QOL evaluation. **METHODS:** 208 patients were studied: 74 had rheumatoid arthritis (RA), 57 connective tissue diseases (CTD), 35 osteoarthritis (OA), 18 non-rheumatoid arthritides (NRA), and 24 miscellaneous diseases (MISC). QOL was evaluated by a specific instrument -the HAQ FDI, 2 generic instruments -the MOS SF-36 and the EQ-5D, and by an individual QOL measure, the Patient Generated Index (PGI). Statistical methods included simple correlations (Pearson and Spearman), ANOVA to analyze differences between groups, and stepwise multiple regression to individuate the best predictors of PGI scores. **RESULTS:** PGI average score was lower in males than in females (3.4 ± 0.2 vs 4.3 ± 0.4 , $p = 0.02$). No correlation was found with age, marital status, BMI, and education ($p > 0.05$). The score was significantly different between disease groups (RA: 3.1 ± 0.2 ; CTD 4.2 ± 0.3 ; NRA: 2.5 ± 0.5 ; OA 3.8 ± 0.3 ; MISC: 4.0 ± 0.4 , sex-adjusted values, $p = 0.006$). PGI score was correlated with erythrocyte sedimentation rate, limitedly to inflammatory disorders ($N = 159$, $p = 0.009$), but not with disease duration and with the presence of fibromyalgia. Significant correlations ($p < 0.01$) were found with all SF-36 scales (r from 0.24 for general health perception, to 0.46 for physical functioning), all EQ-5D scales (r from 0.20 for anxiety/depression to 0.45 for usual activities), and to FDI ($r = 0.41$). However, in the regression model, which included also patient's assessment of disease and pain, and physician's assessment of disease, only EQ-5D usual activities scale, and SF-36 physical functioning, physical role, and mental health scales were retained ($R^2 = 0.299$). **CONCLUSION:** individual quality of life in rheumatic diseases, as assessed by PGI, is not reliably predicted by usual clinical and health status data.

Abstract 1764**VALIDITY OF THE BRIEF PAIN INVENTORY FOR USE IN CLINICAL TRIALS OF NONMALIGNANT PAIN THERAPY**

Susan D. Keller, Carla M. Bann, Sheri Dodd, Health and Social Policy Division, Research Triangle Institute, Research Triangle Park, NC

Many attributes of the Brief Pain Inventory (BPI) make it attractive for use in multinational clinical trials of pain therapy including ease of administration, broadly applicable content, and the availability of translations. Yet its widespread use is inhibited by lack of documentation regarding its reliability and validity. Originally developed for use in cancer patients, there are few published empirical studies of its psychometric properties outside of that population. We sought to assess the validity of the BPI as an outcome measure in patients with arthritis and low back pain by evaluating its relationship to widely-used generic and condition-specific pain measures. The BPI along with the SF-36's Bodily Pain scale (SF-36 BP), the Chronic Pain Grade (CPG), the Roland Disability Questionnaire (RDQ), and the Health Assessment Questionnaire (HAQ) was self-administered to 130 patients in the clinic. Internal consistency reliability coefficients for the BPI Severity and Interference scales were high ($\alpha=0.86$ and 0.94 , respectively) and comparable to those found for the SF-36 BP ($\alpha=0.83$), the RDQ ($\alpha=0.93$) and the HAQ (α ranged from 0.68 to 0.86 across the eight scales). Supporting its interpretation as a generic measure of pain, the BPI scales demonstrated similar patterns of relationships to generic and condition-specific pain and general health scales as did the SF-36 BP. In 43 of 60 comparisons, the magnitude of the correlation between the BPI scales and other generic and condition-specific measures was higher than that of the SF-36 BP. The BPI Severity and Interference scale scores significantly discriminated among groups of patients stratified according to their scores on the CPG, the RDQ or the HAQ as did the SF-36 BP. With one exception, F statistics for the BPI scales were larger than those for the SF-36 BP. Results of this study support the validity of the BPI for use in clinical trials of therapy for patients with arthritis and low back pain.

Abstract 1461**QUALITY OF LIFE OF PATIENTS WITH LOW BACK PAIN BEFORE AND AFTER MICRODISCECTOMY**

Max Leibetseder, Verena K. Greimel, Wolfgang P. Piotrowski, Regine Klinger, Martin A. Krombholz-Reindl, Gabriele Kohlboeck, Richard Neuhofer, Alfred Schmid, Department of Social Medical Service, Government of Salzburg, Salzburg, Salzburg County, Austria

The purpose of this study was to assess and compare the effect on health-related quality of life of patients with low back pain before and after microdiscectomy (lumbar disc herniation L4/L5 and / or L5/S1). Forty seven consecutive patients with low back pain, scheduled for surgery, were included in the study. Health-related quality of life was assessed using the German version of the Medical Outcomes Study, a 36-Item Short Form Health Survey (SF-36). Depression was measured using the General Depression Scale (ADS-L) a modification of the depression inventory of the Center of Epidemiological Studies-Depression (CES-D). Intensity of pain was scored on a 6 point numeric rating scale. Questionnaires were administered one week before and three weeks after surgery, all are self-rated assessments. Statistical analyses showed that quality of life increased after surgery in physical functioning, bodily pain, vitality and mental health. Higher depressed patients showed higher scores in mental health after surgery than lower depressed patients. Patients with higher intensity of pain showed higher reduction of pain after surgery than patients with lower intensity of pain. The higher the intensity of pain and/or depression before surgery the greater the improvement in the corresponding dimensions after surgery. This result is not consistent with other studies on depressed persons. One possible explanation is that depressive mood before surgery was caused by pain and the improvement of mood after surgery can be understood as reaction of short relief. Further research is needed to prove this hypothesis.

Abstract 1162**THE USE OF TRANSITIONAL VERSUS SERIAL-STATE MEASURES AS INDICATORS OF IMPROVEMENT IN STUDIES OF RESPONSIVENESS**

Dorcas E. Beaton, James G. Wright, Sheilah Hogg-Johnson, Claire Bombardier, Department of Research, Institute for Work & Health, Toronto, ON, Canada

Introduction: External markers provide the construct for change in anchor-based studies of responsiveness. Two formats can be used: a transitional question (rating the change), or an arithmetic difference in two state measures (state before minus after). The purpose of this study was to compare two different formats of the external marker. Methods: 200 patients undergoing care of an upper-limb musculoskeletal disorder completed questionnaire packages before and 12 weeks after treatment. The questionnaire included the DASH (disabilities of the arm, shoulder and hand), the SF-36, and the EQ-5D. Ratings of change in the attributes of overall problem, pain and function were gathered using two formats: serial state measures (SSM) (7-10 point scales) or transitional indices(TM) (0-10, 5 = no change). The pattern of responses for each attribute were contrasted with the two formats. Responsiveness (standardized response mean= SRM) of each measure was estimated using the six markers as indicators of improvement and the ratio of SRM SQUARED (SSM/TM) to compare results across formats. Correlations (Spearman) between external markers and change in the DASH were calculated. Results: 172 subjects completed both packages (86%). Patients reported more improvement using the transitional measures. Estimates of responsiveness also varied between formats with the ratio of the responsiveness (square of (SRM-SSM/SRM-TM)) ranging between 1.53 - 1.70 (DASH), 0.89 - 3.42 (SF-36), and 1.43 - 2.36 (EQ-5D). Correlations with change in the DASH and overall rating, pain and function were 0.43, 0.37 and 0.39 (TM), and 0.60, 0.62 and 0.63 using SSM. Correlations between SSM and TM for the same attribute were between 0.39 and 0.42. Conclusions: Differences in responsiveness of questionnaires can be attributed, in part, to the format of the external marker. The results should be considered when trying to interpret the clinical relevance of studies of responsiveness.

Abstract 1695**EARLY IMPACT OF CHEMOTHERAPY ON QUALITY OF LIFE OF MALIGNANT LYMPHOMA PATIENTS**

A. A. Novik, T. I. Ionova, A. S. Povzun, A. V. Kishtovich, T. N. Perekatova, T. P. Nikitiina, Military Medical Academy, Multinational Center of Quality of Life Research, St-Petersburg, Russian Federation

Chemotherapy (CT) has its impact on quality of life (QoL) of cancer patients. The purpose of the study was to identify the effect of the first course of conventional CT on QoL of malignant lymphoma patients. Methods: New patients with aggressive non-Hodgkin's lymphoma (AL), indolent non-Hodgkin's lymphoma (IL) and Hodgkin's disease (HD) were treated by 6 or 8 courses of CHOP, COP and COPP-ABV correspondingly. EORTC QLQ C-30 was administered at base line and before the second course of CT. 72 patients - 34 males (mean age 41,6) and 38 females (mean age 44,1) were studied (29-AL, 14-IL and 29-HD patients). Results: In AL patients there was observed the improvement of patient's functioning and significant increase of general QoL (41,37 vs 51,14) after the first course of CT as compared with base line. At the same time there was the increase of nausea and vomiting (9,77 vs. 24,71). The first course of COP in IL patients led to the improvement of functioning without the increase of symptoms. There was observed significant improvement in role (71,42 vs. 79,76) and emotional (64,88 vs. 76,78) functioning as compared with base line. In HD patients COPP-ABV had negative influence on cognitive functioning (87,35 vs 77,58). There was also observed the nausea and vomiting (10,91 vs. 20,11 and dyspnea (17,24 vs. 31,03) deterioration in HD patients. Conclusions: Conventional CT has the early impact on QoL of malignant lymphoma patients. It depends on the type of lymphoma and CT regimen. In IL patients there was the improvement of functioning without negative impact on symptoms. There was toxicity manifestation and deterioration of cognitive functioning in HD patients. In all patients improvement of general QoL was achieved in spite of the fact of increasing nausea and vomiting.

Abstract 1582**THE USE OF A PATIENT SYMPTOM DIARY TO ASSESS PALLIATION IN LUNG CANCER PATIENTS: AGREEMENT WITH PHYSICIAN ASSESSMENT AND EORTC QLQ SCORES**

Michael D. Brundage, Benny Zee, Michael Palmer, Andrea Bezjak, Peter Dixon, Joseph Pater, Department of Oncology, Queen's University, Kingston, ON, Canada

The purpose of the study was: (a) to assess the extent of concordance on the symptom deemed by the oncologist as the indication for radiotherapy ("index symptom") and that scored the patient as the most distressing, and (b) to compare palliation as measured by a daily diary to that measured by the EORTC QLQ C30 (QLQ). Data were collected in a phase III trial of thoracic radiotherapy for lung cancer (first 88 patients). Patients rated nine symptoms each on a 4-point scale at baseline and then daily until the first follow-up visit at 5 weeks (diary). Patients completed the QLQ at baseline and at 5 weeks. For diary scores, palliation was measured as the change between index symptom baseline scores and the mean of scores for that symptom during the 5th week; for QLQ palliation was the change between baseline and follow-up on each relevant domain (dyspnoea, pain, fatigue, loss of appetite, nausea, vomiting). Oncologists identified the index symptom as pain (25/88, 28%), cough (25%), dyspnoea (23%), fatigue (11%), haemoptysis (10%) or loss of appetite (2%). The index symptom was also rated as the most distressing by 67 (76%) patients ($Kappa=0.7$; $95\%CI:0.6-0.8$); 52 of the 67, however, reported an additional equally distressing symptom. Fatigue and cough accounted for 16/21 discordant cases. Forty-four patients had index symptoms that were measured by both the diary and QLQ domains; the percentage agreement between the QLQ and the diary scores on the identification of palliation was 21/44 patients (48%, in whom symptoms improved in 14 and did not change in 7). The instruments disagreed for the remaining 23 patients: the diaries suggested better palliation than did the QLQ for 16 patients and worse for 7. We conclude that patient self-recorded daily symptom scores give useful information about baseline symptom status and about response to treatment, and often yield different information than physician assessment or QOL scores.

Abstract 1204**THE ROLE OF QUALITY OF LIFE (QL) CONSIDERATIONS IN TREATMENT DECISIONS REGARDING PALLIATIVE CHEMOTHERAPY**

Detmar SB, Wever LDV, Tuijthof van C, Muller MJ, Schornagel JH, Aaronson NK (The Netherlands Cancer Institute, Amsterdam, The Netherlands)

Background: Although improving QL is an important goal of palliative treatment, very little is known about the importance of QL considerations on treatment decisions. The aim of this study is to describe the relative role of tumor response, hematological toxicity, and QL factors on the decision to adjust or stop chemotherapy during a palliative treatment period. **Methods:** The study sample consisted of 217 outpatients receiving palliative chemotherapy. After the second chemotherapy cycle 4 consecutive consultations were audiotaped and content-analyzed to determine how often and for which reasons treatment alterations were made. Additionally, physicians rated their patients' QL during each visit using the COOP/WONCA charts, and data on tumor response and toxicity were obtained from the medical records. **Results:** Decisions to change or stop the treatment were made in 114 cases, primarily due to tumor progression (53) and hematological treatment toxicity (35). In fact, tumor progression and serious toxicity always led to alterations in treatment, regardless of QL considerations. In 26 cases, QL factors played a prominent role in the decision to change or stop the treatment. For those patients for whom QL considerations were an important reason for altering the treatment, previous use of chemotherapy and the perceived importance for patients of the effect of chemotherapy on daily life were significantly related to this decision. **Conclusions:** In contrast to what is often suggested in the literature, QL factors appear to play a significant role in decisions about palliative chemotherapy treatment in only a minority of cases. Rather, tumor progression and hematological toxicity are the most important reasons for changing or discontinuing treatment.

Abstract 1167**COMPARISON OF THREE QUALITY-OF-LIFE INSTRUMENTS FOR CANCER PATIENTS : FACT-G , EORTC QLQ-C30 AND FLIC**

Thierry Conroy, Mariette Mercier, Simon Schraub, Jacques Bonnetterre, Elisabeth Luporsi, Jean-Louis Lefebvre, Michel Lapeyre, Marc Puyraveau, Department of Medical Oncology, Centre Alexis Vautrin, Vandoeuvre-lès-Nancy Cedex, France

Between 10/1996 and 6/1999, 310 patients with breast cancer (BC;n=163), head and neck cancer (HNC;n=87) or colorectal cancer (CRC;n=60) completed FACT-G version 3 (french version) and EORTC QLQ-C30 at the same time. BC and CRC pts also completed the FLIC questionnaire (q). HNC pts received 2 specific HN modules (FACT H&N ,HN Besançon q) and completed the 4 q during the 1st and last week of radiotherapy (RT). BC pts filled q at day 8 of 2 consecutive cycles of adjuvant chemotherapy (CT). CRC pts completed q once during cycle 2 or 3 of CT. All q were given simultaneously. FACT-G has to be completed first, then the other q according to a randomized order. 37.4 % of BC and CRC pts preferred QLQ-C30 and 21.5 % FACT-G ($p=0.0025$). 16.8 % preferred FLIC. Most HNC pts had no preference. For CRC and HNC pts, time to complete FACT-G (11') was longer than for the other q (8'). Items of FACT-G were more frequently quoted as confusing (1.5 %) or upsetting (0.6 %) vs 0.36 % and 0.03 % of QLQ-C30 items and 1 % and 0.17% of FLIC. Data were missing for 5.23 % of FACT items, 1.76 % of QLQ-C30 and 0.3 % of FLIC. Construct and content validity, internal consistency of french version of FACT-G v3 was demonstrated but rewording of items 13 and 21 seemed necessary. Concordant validity between FACT-G and QLQ-C30 was good, except for social domain. Reproducibility, tested in BC pts, was good. QoL scores were lower in pts with BC and in HNC living alone. Responsiveness in HNC was better for specific modules (changes :22.3 % for FACT H&N, 25.3 % for HN Besançon q vs 4.5 % for FACT-G). In case of toxicity (WHO grade) in these pts, FACT-G scores' variations were greater than those of EORTC global QoL. In conclusion, FACT-G was validated in french language, but EORTC QLQ-C30 was preferred by pts. Addition of a specific module increased sensitivity to changes in HNC pts. Supported by french government's PHRC.

Abstract 1551**QUALITY OF LIFE IN LONG-TERM SURVIVORS OF LUNG CANCER**

Linda Sarna, Geraldine Padilla, School of Nursing, University of California, Los Angeles, Los Angeles, CA

Lung cancer will impact over 1 million Americans in the next decade. The majority of quality of life (QOL) research has focused on persons with advanced lung cancer. Little is known about the QOL of the growing number of long-term survivors. The purpose of this investigation was to describe the QOL of 150 minimum 5-year disease-free survivors of non-small cell lung cancer. In a preliminary analysis of the first 100 participants, the sample is: 56% female, 82% white, 46% married, and the average age is 70 years (range 39 - 89 yrs). The time since diagnosis ranges from 5 - 21 years (mean = 11 yrs), 59% had adenocarcinoma; and had an average of 2 comorbid conditions. Seven percent were current smokers, 77% former smokers, and 6% never smokers. Pulmonary function varied widely (range 27% - 128% FEV1). A cross-sectional survey approach was used to assess QOL (Ferrell QOL-Survivor tool & the SF-36) and explore variables associated with QOL. The mean scores for Overall QOL, Psychological, Physical, and Social dimensions of the Ferrell tool were as high or higher than previously reported for other cancer populations. Older age and better pulmonary function were associated with better overall QOL; depression (CES-D over 16) and number of comorbid conditions were associated with worse QOL. Higher Physical QOL scores were associated with better pulmonary function, fewer comorbid conditions, and lower CES-D scores. The average SF-36 Physical Well-being dimension normed for cancer was in the 75th percentile. The Psychological dimension of the survivor tool was positively related to older age and negatively with comorbid conditions; the cancer-normed Mental Well-Being score of the SF-36 was at the 50th percentile. The Social QOL dimension was related to age, number of comorbid conditions, and depression. The Spiritual dimension score was significantly higher among women. Lung cancer survivors enjoy higher QOL than that reported for other survivors of cancer. Comorbid disease, not older age, was associated with poorer QOL. (Funded by NCI RO1 CA78997)

Abstract 1700**POSITIVE EPOETIN ALFA EFFECT ON HEALTH-RELATED QUALITY OF LIFE (HRQOL) IN ANEMIC CANCER PATIENTS (PTS) RECEIVING CHEMOTHERAPY**

Martin Zagari, Dennis Gagnon, Department of Health Economics, Johnson & Johnson, Raritan, NJ

Fatigue, often associated with chemotherapy-associated anemia, is common in cancer pts. Several clinical trials have demonstrated that treatment of anemia may lead to improvement not only in levels of fatigue, but in more general measures of HrQOL as well. The effect of epoetin alfa on HrQOL was assessed in a number of trials including 2 randomized, double-blind trials enrolling 349 and 213 anemic cancer pts evaluable for HrQOL and 2 open-label, uncontrolled community trials enrolling >4000 pts total. All trials included 3 single-item Cancer Linear Analog Scales (CLAS, also known as Linear Analog Scale Assessment [LASA]) for Energy (E), Daily Activities, and Overall QOL (OQOL). One randomized and one open-label study contained the multi-item Functional Assessment of Cancer Therapy (FACT) General (G) and Fatigue (F) scales, and one randomized trial contained the multi-item Short Form-36 (SF-36). The FACT-An and CLAS, known to be sensitive to change in hemoglobin (Hb), were particularly suitable for use in these trials. The more general SF-36 assessed HrQOL in generic domains. In the largest randomized trial, epoetin alfa pts experienced significant improvements in the 3 CLAS measures and in the FACT G and F scales ($P < .05$). In addition there were trends in improvement for the SF-36 Physical and Mental Component Summary scores ($P < .1$). Pts treated with epoetin alfa in the smaller randomized trial showed significant improvement in OQOL ($P < .01$) and a trend towards significant improvement in E ($P < .1$). The magnitude of change in all CLAS scores was comparable to that seen in the larger study. In the large community-based trials, epoetin alfa pts experienced significant improvements from baseline for the CLAS and FACT measures ($P < .001$). Increase in Hb was associated with improved HrQOL in all trials. In summary, improvement in HrQOL is associated with epoetin alfa treatment, as mediated by improvement in Hb. These relationships have been replicated in other trials using various HrQOL instruments.

Abstract 1697**THE SIGNIFICANCE OF MISSING DATA IN CLINICAL TRIALS**

Diane L. Fairclough, Dennis Gagnon, Center for Research Methodology and Biometry, AMC Cancer Research Center, Denver, CO

The significance of missing data in clinical trials depends on the underlying mechanism, the research question, the population of inference, and the analytic model. A series of trials (Phases III and IV) of epoetin alfa for the treatment of chemotherapy-induced anemia illustrates the need to assess the significance of missing data on study conclusions. In the health-related quality of life (HrQOL) analyses of a double-blind, randomized Phase III trial of epoetin alfa vs standard treatment of anemia, the presence of nonignorable missing data was hypothesized with the main concern being the potential impact of missing observations on the tests of between-group differences in HrQOL. Using a joint mixed-effects model for HrQOL and time to treatment discontinuation, the moderate correlations ($\rho = 0.46-0.61$) of each individual's change in HrQOL over time with the time of treatment indicated a nonignorable process. Although the estimates of change over time were sensitive to the missing data assumptions, the between-group results of the joint mixed-effects model were comparable to the estimates obtained assuming an ignorable process. In HrQOL analyses of an open-label, nonrandomized, community-based Phase IV trial, the research question was the relationship between hemoglobin (Hb) and HrQOL. Using a 2-stage procedure introduced by Heckman ("Heckit"), there was evidence of a covariate-dependent dropout (changes in Hb, platinum-based therapy, initial HrQOL, and survival). In the second stage, however, missing data appeared to be ignorable, conditional on covariates (eg, changes in Hb, pain, and nausea and disease status) and observed HrQOL. In a second open-label trial that did not include the same wealth of covariate information, there was a statistically significant (probably nonignorable) association between HrQOL and dropout. In summary, the treatment of missing data and the determination of whether it is ignorable depends on the research question and the covariates available for analysis.

Abstract 1688**MULTILINGUAL HEALTH-RELATED QUALITY OF LIFE (HRQOL) ASSESSMENT AND POOLING OF DATA FROM INTERNATIONAL CLINICAL TRIALS**

David F. Cella, Chih-Hung Chang, Dennis Gagnon, Brian Bresnahan, Martin Zagari, Institute of Health Services Research Policy, Evanston Northwestern Healthcare, Evanston, IL

Clinical trials including HrQOL endpoints are increasingly conducted in many cultures and languages. Pooling of data across countries and cultures within such trials requires examination of language-dependent variations in HrQOL responses and the effect of pooling on the results of trials. HrQOL data from a multinational, randomized, placebo-controlled clinical trial to assess the effect of epoetin alfa on cancer-related anemia were analyzed to evaluate the measurement characteristics of 6 language versions of the Functional Assessment of Cancer Therapy-Anemia (FACT-An): Dutch, French, German, Italian, Portuguese, and English. Classical test theory and item response theory were used to test for internal consistency and differential item functioning (DIF), respectively, across the different language versions. Traditional measures of reliability showed that internal consistency was high across all languages for both of the clinical trial's FACT-An-derived primary endpoints: FACT-G total (range: .83-.91) and the Fatigue subscale (range: .89-.92). Using English as the reference language, Rasch rating scale analysis was used to perform pairwise comparisons for each of the 5 target languages. The analysis generated relative item location hierarchies for each language with a criterion of >0.5 logit used to define DIF. Results revealed 21 of 200 (10.5%) cases of DIF for the FACT-G and Fatigue items. An 89.5% item location invariance justifies pooling of data for the primary endpoint analyses. All-item endpoint analysis can then be compared to "invariant-item-only" endpoint analysis to support the decision. Future challenges include defining conditions (eg, degree of DIF) under which language or culture variation may be too excessive to support pooling, and how to interpret multicultural trial results when the same questions perform differently across languages or cultures.

Abstract 1708**ASSESSING THE CLINICAL SIGNIFICANCE OF CHANGES IN HEALTH-RELATED QUALITY OF LIFE (HRQOL) SCORES**

Donald L. Patrick, Dennis Gagnon, Martin Zagari, University of Washington, Seattle, WA

Despite recent advances in HrQOL measurement, experience with most instruments remains limited. Alternatives for assigning clinical meaning are being explored to help interpret changes in HrQOL scores. A recently completed study in which epoetin alfa was compared to standard management of anemia used 7 primary scales to measure HrQOL: CLAS¹, also known as LASA², Energy (E), Daily Activities (DA), and Overall QOL (OQOL); FACT³ General (G) and Fatigue (F); and SF-36⁴ PCS⁵ and MCS⁶. Three methods for interpretation were explored. (1) HrQOL scores were interpreted directly from observed change in hemoglobin (Hb). A mean difference in Hb change of 1.7 g/dL favoring epoetin alfa corresponded to HrQOL changes of: E (13.87), DA (13.50), OQOL (10.76), G (5.73), F (5.15), PCS (2.30), and MCS (2.39). (2) Change in HrQOL was compared to "minimal clinically important" differences (MCID) in Hb to help establish a level of HrQOL change that may be considered clinically significant. Setting this threshold at 1 g/dL (mean change between groups was 2.78 g/dL) yielded the following MCIDs: E (9.59), DA (9.22), OQOL (10.19), G (2.25), F (4.22), PCS (2.74), and MCS (-0.89). Other thresholds were explored. (3) Regression analyses determined that a 1-g/dL improvement in Hb change (eg, comparing a 1-g/dL change to a 2-g/dL change) corresponded to HrQOL improvement of: E (4.34), DA (5.00), OQOL (4.50), G (1.86), F (1.60), PCS (0.86), and MCS (0.73). These analyses suggest HrQOL changes in these instruments that have been observed in clinical trials were clinically significant based on Hb changes that are used to guide clinical practice. These analyses can be performed retrospectively, but ideally they could be applied *a priori* to evaluate changes in HrQOL that would be considered "significant."¹Cancer Linear Analog Scales; ²Linear Analog Scale Assessment; ³Functional Assessment of Cancer Therapy; ⁴Medical Outcomes Study-Short Form 36; ⁵Physical and ⁶Mental Component Summary

Abstract 1122

MEASUREMENT OF TREATMENT SATISFACTION

Catherine Copley-Merriman, Andrew Wang, Outcomes Research Department, Parke-Davis Pharmaceutical Research, Ann Arbor, MI

Treatment Satisfaction has become a new area of interest in outcomes research, but the published literature focusing on the measurement of treatment satisfaction is limited. **OBJECTIVES:** The purpose of this review is to describe the construct of measures and common domains, and show examples of surveys found in the literature. **METHODS:** English-language studies published between September 1989 to September 1999 were identified through a MEDLINE search, and a few additional articles were added after that timepoint. **RESULTS:** The present review indicates a lack of a well-supported conceptual model of treatment satisfaction. Instruments were often ad hoc questions added on to other measures in the studies. Likert-scaling is the most frequently used method. Common domains include satisfaction with outcomes (including efficacy, speed to onset of desired effect, and feeling about side effects), process (ease and convenience of therapy, and cost of therapy) and overall satisfaction (including desire to continue therapy, degree treatment met expectations and recommendation to others). The limited number of validated disease-specific measures for treatment satisfaction include The Diabetes Treatment Satisfaction Questionnaire, The Seattle Angina Questionnaire, the Pain Service Satisfaction Test and the Erectile Dysfunction Inventory of Treatment Satisfaction. A few studies have shown that patient evaluation of different treatments (or treatment regimens) using unvalidated treatment satisfaction questions can result in statistically significant differences. The studies of treatment satisfaction consistently show that it is measuring something different from HRQL, but it should be evaluated in conjunction with other study outcomes. **CONCLUSION:** Treatment satisfaction is rapidly becoming an area of high interest for outcomes research, but more work is needed to develop a standard construct and validation methods.

Abstract 1361

MEASURING THE INFLUENCE OF PARENTAL SATISFACTION WITH THERAPY ON CHILD ADHERENCE TO ASTHMA MEDICATION: AN INTERNATIONAL VALIDATION

Linda M. Nelsen, Nancy C. Santanello, Steven Bird, Joris Menten, Jonathan M. Edelman, Jonathan A. Leff, Department of Epidemiology, Merck & Co., Inc., West Point, PA

In prescribing therapies for asthmatic children, understanding elements of satisfaction of the parent, who primarily supervises therapy, may lead to better adherence and improved asthma outcomes. The influence of parental satisfaction with therapy on the child's medication adherence and on asthma outcomes were evaluated with data from 2 identical crossover trials of asthma therapy in children ages 6-11 (I:US n=326; II:ex-US n=266). After each 4 week treatment period, parents completed a medication satisfaction questionnaire. Based on daily diaries, adherence was measured as the percent of days in full compliance with the prescribed regimen, and outcomes as the occurrence of asthma exacerbations and average daily puffs of beta agonist (BA). A parent satisfaction questionnaire included 6 items including one on overall satisfaction. At baseline, the mean age was 8.8±1.6 years, 35% were female and FEV1 was 74% of predicted. Factor analysis confirmed 2 domains for the parent satisfaction questionnaire: parent bother (BOTHER) and parent perception of symptom control (CONTROL). In both studies, Cronbach's alpha was high for both domains (alpha>0.78) and both domains were correlated with overall parent and child satisfaction (r=0.63-0.75). Measurement characteristics, including construct validity, discriminant validity, and responsiveness were demonstrated for both domains, with similar results for both the US and ex-US studies. Children of parents reporting higher satisfaction with therapy had significantly more adherent days (I: 90% vs 78%, p=0.002; II: 88% vs 78%, p=0.006 of days). Adherent children had fewer exacerbations (I: 2.7% vs 13.6%, p=0.001; II: 3.3% vs 18.9%, p=0.0003). In study II, adherent children had significantly lower use of BA. Thus, parent satisfaction with therapy for their asthmatic children was associated with higher adherence, and improved outcomes of therapy.

Abstract 1696

RELATIONSHIP OF TREATMENT SATISFACTION TO OTHER OUTCOME MEASURES IN MIGRAINE

Nancy C. Santanello, Glenn Davies, Department of Epidemiology, Merck Research Laboratories, West Point, PA

Satisfaction with medication treatment is a multifaceted outcome that may be based on patient characteristics and medication attributes. Drug factors may include ease of administration, dosing, efficacy, tolerability, and other drug related factors. Understanding the relationship between treatment satisfaction and other outcomes may lead to better treatments and consequent improved health outcomes. We examined the relationship between treatment satisfaction and other outcomes using data from 1,506 migraineurs randomized to 2 placebo controlled trials at 47 centers in 21 countries. Satisfaction with migraine treatment was assessed 2 hours after treatment with an overall satisfaction question (7-point scale: 'completely satisfied' to 'completely dissatisfied'). Satisfaction was related to complete pain relief, speed of relief, return to functioning, and relief of associated symptoms of nausea, vomiting, photo- and phonophobia. High rates of satisfaction were reported in the pain free group (>90%), low rates in the group with moderate to severe pain (<10%) at 2 hours after treatment; the group with mild pain at 2 hours experienced intermediate rates. Logistic regression models for mild pain showed the probability of being satisfied was about 80% if baseline pain was severe, associated symptoms were absent at 2 hours and mild pain relief achieved within the first hour of taking therapy. Patients were satisfied in approximately 75% of attacks where the patient reported no associated symptoms compared to 30-40% in attacks where patients had at least one associated symptom. Migraine-specific quality of life (MQoL) measured at the end of the migraine attack was significantly better in patients who reported satisfaction at 2 hours (p<0.0001). In summary, the most important migraine drug attribute was efficacy (fast, complete relief) in determining short-term patient satisfaction with treatment. Satisfaction with migraine therapy at 2 hours after dosing was related to improved MQoL over the entire attack period.

Abstract 1779

RELATIONSHIP AMONG HRQL, SYMPTOM SEVERITY, SATISFACTION WITH HEALTH AND TREATMENT SATISFACTION IN CHRONIC CONSTIPATION (CC)

Margaret L. Rothman, Dominique Dubois, Carmela Janagap, Dennis A. Revicki, Department of Health Economics, J&J, Titusville, NJ

Constipation is a prevalent disorder. Diagnosis is dependent on symptoms and in its more severe manifestation can result in disruptions of HRQL. Treatment of CC has been limited, but new therapies are currently undergoing evaluation in clinical trials (CT). It is increasingly recognized that subject experiences offer unique input beyond what may be obtained by traditional clinical endpoints. Such experience may be categorized into several constructs including HRQL, symptom severity (SX), satisfaction with health or functioning (SF), and satisfaction with treatment (SaTx). Examination of correlations among these variables suggests a relationship, but the structure of this relationship has not been examined. The purpose of this study is to elaborate the relationship among these constructs in a sample of subjects with CC. HRQL, SX, and SF and SaTx were assessed in the context of a CT to evaluate prucalopride versus placebo in the treatment of CC (n=570). These constructs were measured using the PAC-SYM and PAC-QOL, recently developed constipation specific, subject-report scales. Relationships among the constructs were assessed using structural equation modeling. Results indicated that the latent variables HRQL and SX are highly correlated (r=.83). Comparison of hierarchical models assessing the direct and indirect effects of HRQL and SX on SaTx suggests that their impact is mediated by SF (difference in chi square/df = 20.1/2, p <.01). HRQL and SX account for 51% of the variance in SF. The findings suggest that HRQL and SX have an indirect effect on SaTx that is mediated by SF. The construct SF appears to provide unique information in this sample of subjects with CC and should be considered for inclusion in CT.

Abstract 1017**COMPARISON OF THE HEALTH-RELATED QUALITY OF LIFE OF CANCER AND NON-CANCER PATIENTS**

Frank Baker, Vice President for Behavioral Research, American Cancer Society, Atlanta, GA

Data from the 1998 Health Care Financing Administration's Medicare Health Outcomes Survey of patients enrolled in managed care through Medicare were analyzed to compare the health-related quality of life of cancer survivors with age-matched non-cancer patients. A group of 23,137 respondents who answered positively when asked if a doctor had ever told them that they had any cancer were compared on Medical Outcomes Study Short Form -36 (MOS SF-36) scores to an age-matched sample who indicated that they had not been diagnosed with cancer. Cancer patients had significantly poorer scores on all eight of the subscales of the MOS SF-36 as well as on the Physical and Mental Component summary measures. Comparisons of the eight SF-36 subscale scores of the cancer survivors with those of the non-cancer patients and with general population SF-36 norms for people of comparable age are being utilized to help set specific quality-of-life goals for the American Cancer Society's efforts in improving cancer survivor quality of life by the year 2015.

Abstract 1018**ASSESSING THE HEALTH STATUS OF CHINESE AND SPANISH-SPEAKING BENEFICIARIES IN MEDICARE MANAGED CARE**

Barbara L. Gandek, Xinhua S. Ren, Hernan Quinones, John E. Ware, Jr., Health Assessment Lab, Boston, MA

To expand the opportunity for participation in the Medicare Managed Care Health Outcomes Survey (MHOS), the MHOS has been translated into Spanish and Chinese. In 1999, 203 and 669 beneficiaries completed the Spanish and Chinese MHOS, respectively. Tests of assumptions underlying the scoring and interpretation of SF-36 scales generally were met. Internal consistency reliability (Cronbach's alpha) ranged from 0.84 to 0.94 in the Spanish and 0.76 to 0.90 in the Chinese data. Respondents who reported 1 or more medical conditions scored lower on all SF-36 measures, supporting construct validity. The pattern of scale-factor correlations for Chinese and Spanish respondents differed from English-language groups, using a 2 factor solution with varimax rotation, particularly for less educated Spanish and Chinese respondents. These differences were taken into account in interpreting results. Scores for Spanish and Chinese beneficiaries were compared to a random sample of English-language MHOS data (n=5207). Spanish-language respondents were younger (68.3 for Spanish, 70.4 for Chinese, 73.3 for English) and reported more medical conditions on average (2.7 out of 13 conditions for Spanish, 1.3 for Chinese, 2.1 for English). Respondents who completed the MHOS in Spanish also were younger and less healthy than Hispanic respondents who answered the MHOS in English. Even after adjustment for age, gender, education, and medical conditions, there were substantial differences in scores for the 8 SF-36 scales and 2 summary measures across language groups. Spanish beneficiaries scored significantly lower than Chinese respondents on 9 out of 10 measures and significantly lower than the English-language group on 8 out of 10 measures. Chinese beneficiaries generally scored significantly higher than English respondents on physical measures, but significantly lower on mental measures. These findings, at least in part, reflect true differences in health, although cultural and linguistic differences warrant further study.

Abstract 1019**ELDERLY WOMEN IN MANAGED CARE: IMPACT OF RACE/ETHNICITY, EDUCATION, AND INCOME ON HRQOL**

Arlene S. Bierman, C S. Haffer, Y Hwang, J Mandelblatt, Center for Outcomes and Effectiveness Research, Agency for Healthcare Research and Quality, Rockville, MD

Gender differences in disease prevalence and disability influence the HRQOL of elderly women. Among women sociodemographic characteristics are associated with differences in health and functional status. Because of sample size limitations, prior studies have been unable to adequately assess the independent effects of, and interactions between education, race/ethnicity (R/E), and income as measures of socioeconomic status (SES) and health status in older women. We assess these factors using data from the 1999 Medicare Health Outcomes Study. The survey randomly sampled 1000 enrollees from each Medicare managed care plan in the US and had a response rate of 66%. Our analysis includes 87,220 women age 65 and older including 5907 African American (AA), 4376 Latina, and 1394 Asian women. To assess HRQOL we used the two SF-36 summary scores (PCS and MCS) and the 8 subscales of the SF-36. The contribution of comorbidity, measured by a count of self-reported chronic conditions, was assessed. In bivariate analysis there were significant differences in SES, comorbidity, and SF-36 scores by R/E. 27% white, 30% Asian, 49% AA, and 55% Latina women reported < HS education. Mean SF-36 scores were highest for Asian and lowest for AA women. In multivariable models age, income, education, and R/E were independently associated with SF-36 summary scores and subscales which was explained in part by greater comorbidity among women with lower SES. In subgroup analysis a gradient of PCS scores with income was observed among women of each R/E group. Low income white and AA women had equally low adjusted PCS scores (34.4), Latina women had somewhat better scores (36.3) and Asian women the highest (39.2). In designing models of health care delivery, and in assessing outcomes and quality of care for elderly women attention needs to be paid to identified differences in health and functional status among different subgroups of socioeconomically disadvantaged elderly women.

Abstract 1020**ASSESSING THE HRQOL OF SOCIOECONOMICALLY DISADVANTAGED ELDERS IN MEDICARE MANAGED CARE**

Samuel C. Haffer, A S. Bierman, Y Hwang, J Mandelblatt, Director, Medicare Health Outcomes Survey, HCFA, Baltimore, MD

As part of its value-based purchasing strategy, the Health Care Financing Administration (HCFA) has undertaken an initiative to measure and improve the health status of Medicare enrollees in managed care plans. Of interest to policy makers are the differential effects of socioeconomic status (SES) on HRQOL. The 1999 Medicare Health Outcomes Survey (HOS) is an annual survey administered to a random sample of 1000 Medicare beneficiaries in each managed care plan who had been continuously enrolled for at least six months. The 1999 cohort had a response rate of 66%. Our analysis includes the 159,950 individuals age 65 and older who responded. Using the two summary scores from the MOS Short Form 36 (SF-36) as the primary measures of physical functional status and emotional well being, and the 8 SF-36 subscales, we compared the health status of beneficiaries across 3 measures of SES: race/ethnicity, education and income. Interactions between these measures were examined. Bivariate analysis showed significant differences in baseline physical and mental health status across all categories of race/ethnicity, education and income. In multivariable regression analysis income, education, and race are independently associated with function. A gradient of SF-36 scores was noted across all levels of income and education with low income and less education associated with lower scores for PCS, MCS, and all SF-36 subscales. For example, after adjusting for age, gender, race, and education persons with low, middle, and high incomes had PCS scores of 36.9, 39.4, and 42.5 respectively. Adjusting for income and education African Americans reported worse and Asians better HRQOL. Future analysis should focus on determining factors which account for these differences. HCFA may wish to explore methods for ensuring that beneficiaries of lower SES have equal access to needed services, that care delivered improves their functional health outcomes, and that resources are allocated to meet their greater needs.