objectives were to 1) develop a simulation model which can simulate QoL and different cost categories from a societal perspective as affected by AS treatments with sequential drugs using the ASAS recommendations at group level; 2) parameterize the model using Dutch data; and 3) analyze simulated cost-effectiveness of treatment strategies with and without anti-TNFs. METHODS: Discrete-event paradigm was selected for model development. Pathways in AS treatments were simulated in conjunction with changes in AS-specific measures. Parameterization was realized using original data from an observational cohort of AS patients and literature where necessary. Frequencies of visits to rheumatologists and hospitals, and having paramedical treatments, informal/formal care, sick leave, and work disability were estimated using logit and Poisson regression models. Health utility was estimated using a proportional-odds model. Demographic variables were also considered in model selections. Two competing strategies in the simulation were 1) three Cox-1 plus two Cox-2 selective nonsteroidal anti-inflammatory drugs, and 2) same as strategy 1 plus Enandocet and Infliximab. The simulation run was for 1000 patients until death. RESULTS: Simulated values of QALYs and different cost categories were in reasonable ranges. Productivity and hospital costs constituted largest shares of the total costs, and were significantly lower in strategy 2 (P < 0.05). The incremental cost per QALY gained of strategy 2 against strategy 1 was $25,165. CONCLUSIONS: The model is flexible and promising for assessing actual costs-effectiveness of different treatment strategies for different societies.

PODIUM SESSION II: IRT AND RACH MODELING IN QUALITY OF LIFE MEASUREMENT

METHODS FOR PSYCHOMETRIC AND CLINICAL EVALUATIONS OF CAT-BASED MEASURES OF DISEASE IMPACT IN CHRONIC KIDNEY DISEASE (CKD)

Lin P1, Ware JE1, Meyer K2, Richardson M3, Bjorner JB4

1Department of Health Services Research, Kaiser Permanente, Oakland, CA, USA; 2Department of Biostatistics, University of California, Berkeley, CA, USA; 3Tufts Medical Center, Boston, MA, USA; 4National Institute for Environmental Health Sciences, Bethesda, MD, USA

OBJECTIVES: To evaluate a new standardized disease-specific PRO impact item bank, the Kidney Disease Impact Scale (KDIS), among adults with chronic kidney disease (CKD). METHODS: A 34-item bank of items measuring disease impact, with standardized content (differing only in disease attribution) and standardized IRT-based parameters from previous studies across therapeutic areas, was administered to 516 CKD patients, along with KDQOL-CKD, KDQOL-Lite, HQOLI (Effects and Burden scales) and a generic (SF-12v2) health survey. All measures were administered in a clinical setting at three time points (baseline, 1 week, and 3 months) to evaluate: 1) reliability (internal consistency and 1-week test-retest); 2) validity in discriminating among CKD groups differing in clinically defined disease severity (nonnalyses stages 3–5, dialysis patients, and transplant patients); and 3) responsiveness (better, same, or worse after 3 months). Real-data CAT simulations were performed, using only the responses to items that would have been selected and asked during a real CAT, to estimate KDIS scores based on the full bank, and 5-, 10-, and 15-item CATs. RESULTS: As hypothesized, CAT-based score distributions were less skewed and estimates were more efficient (covered a wider range and were as reliable with fewer items), in comparison with legacy measures. All KDIS forms (full bank, various lengths of CATs) discriminated among clinically defined groups as well or better than legacy CKD-specific measures and were substantially better than generic (KDQOL) measures. Relationship between the ability (IRT) coefficients for discriminant validity and responsiveness tests for 1-minute (5-item) CAT administration compared favorably with scores based on the entire bank (0.84, KDQOL scores (0.79 for Burden and 0.50 for Effects), and generic measures (RV = 0.00 to 0.59; medium = 0.13). CONCLUSIONS: Disease-specific PROs can be improved to be more practical and more valid, in comparison with legacy CKD-specific and generic PROs. CKD impact measures that perform better psychometrically tend to perform better in empirical tests of clinical validity.

SCALING PROPERTIES OF TWO COMMONLY USED OUTCOME MEASURES IN DERMATOLOGY—THE DERMATOLOGY LIFE QUALITY INDEX (DLQI) AND THE PSORIASIS QUALITY OF LIFE SCALE (PSORIQoL)

Twisk J1, Mekenna S, Crawford S, Doward L

1Grad Research Limited, Manchester, UK

OBJECTIVES: The DLQI is a 10-item general dermatological HRQoL measure used widely with psoriasis patients. The PSORIQoL is a 25-item psoriasis-specific measure that employs the needs-based model of quality of life (QoL). Both measures have been shown to have good psychometric properties using Classical Test Theory. However, their scaling properties have not been fully assessed using Item Response Theory (IRT). This is the purpose of the study. METHODS: Psoriasis patients, recruited from an out-patients clinic, completed the DLQI and PSORIQoL. IRT (Rasch analysis) was used to determine each measure’s overall fit to the Rasch model, individual item fit, targeting of scales to severity of respondents, functioning of response categories, and the presence of Differential Item Functioning (DIF) by age or gender. RESULTS: 146 psoriasis patients were included in the study (male 50%, mean age = 44.2, range = 17–73 years). DLQI: There was overall misfit to the Rasch model (χ² = 39.65, df = 20, P = 0.005). One item misfit the Rasch model and another showed borderline misfit. Four of the 10 items had disordered response thresholds indicating that they did not work in a logical way. Too few DLQI items covered milder levels of HRQoL (meaning that the measure is not sensitive to change in patients with mild disease). DIF by age was found in one item and DIF by gender in another. PSORIQoL: Overall, the measure fit the Rasch model (χ² = 56.45, df = 50, P = 0.247), although one item had a high fit residual suggesting misfit. The response options worked logically and items were well distributed across the QoL measurement range. No DIF by age or gender was found. CONCLUSIONS: Application of Rasch analysis indicated that there were several problems with the scaling properties of the DLQI. In contrast, the PSORIQoL fit the Rasch model and exhibited good measurement properties.

THE USE OF IRT MODELING TO ACCOUNT FOR DIFFERENTIAL ITEM FUNCTIONING IN THE PROQOL-HIV QUESTIONNAIRE

Lalanne C, Duracinsky M, Armstrong AR, Chassany O

Assistance Publique-Hopitaux de Paris, Paris, France

OBJECTIVES: This study aims to provide sample-free estimates of uniform DIF on a newly developed HRQoL questionnaire, PROQOL-HIV, based on IRT. Patient characteristics considered here were gender, country of residence, and ethnicity, given the high prevalence of HIV in immigrant populations. METHODS: We use a two-step operationalistic regression procedure to model sources of covariates, on two constructs extracted from factor analysis: Physical Health and Symptoms (PHS, 14 items), and Emotional distress and Health Concerns (EDHC, 10 items), a subset of N = 503 patients (71% males, median age 43 years) from five countries (Australia, United States, France, Brazil, and Thailand) was used. Person parameter estimates were obtained using a Graded Response Model. Missing responses (≤1%) were imputed using half-rule. RESULTS: The EDHC dimension was more prone to exhibit item biases compared to PHS items, especially when referring to ethnicity. Country-related effects were equally distributed on the two scales. In both cases, however, no DIF effects were found for gender. We reached similar conclusions when comparing the results to those obtained when using sum or rest scores as conditioning ability levels. Correlations between raw scores computed with or without flagged items were above 0.5. Restricting analyses to Western countries, however, yielded smaller hits. Altogether, these results suggest that DIF effects should be of limited impact when reporting scores for these two HRQoL dimensions of the PROQOL-HIV questionnaire. Results are discussed in light of other latent variable models proposed to assess DIF, the MMIC, and logistic mixed effects model. CONCLUSIONS: Using IRT instead of summed scale scores allows the incorporation of person covariates and the study of DIF effects in a probabilistic framework. The presence of DIF effects on biopsychological constructs raises important cross-cultural validity considerations for multi-country multi-ethnic studies, although assessment of DIF impact needs further investigations.

HUNTINGTON QUALITY OF LIFE INTERVIEW (HQOLI)—A NEW HUNTINGTON’S DISEASE-SPECIFIC QUALITY OF LIFE INSTRUMENT


Cétegrit, Cétegrit, Paris, France; NEUROSEARCH, Ballerup, Denmark; Neuroligic Unit CHU, Angers, France; Université Claude Bernard Lyon 1, Lyon, France

OBJECTIVES: No specific health-related quality of life (HRQOL) instrument tool is widely used for Huntington’s disease (HD). In the context of a large European study on HD burden (Euro-HBD), a specific HRQOL tool (HQOLI) was created for patients in early and intermediate stages of HD. We describe the development and psychometric validation of HQOLI. METHODS: After semi-structured interviews with patients, caregivers, and HD specialists, we formed the HQOLI. The tool was validated in French language. The questionnaire consists of 42 items, relating to 11 dimensions: Mobility, Motor Ability, Autonomy 1 and 2, Finance, Occupational Activities, Psychology, Feelings of Guilt, Vitality, Socializing, and Body Image. Face validity and internal and external validity were examined. RESULTS: Among 178 participating patients, 15 did not complete HQOLI. Item response rates range from 0.81 to 0.98, excluding the occupational activities domain. As expected, there is a ceiling effect on several items, suggesting low sensitivity at advanced HD stage. Combat’s alphas are greater than 0.76. HQOLI total score is highly correlated with generic HR-QOL (SF36 and EQ5D) as well as with the specific caregiver HQRO instrument (HDQol-C). Correlations were –0.73 (P < 0.01) with EQ5D utility, –0.76 (P < 0.01) with “Physical Functioning,” and –0.74 (P < 0.01) with “Mental Health.” Voluntary movement, fall/balance, and dystonia symptoms contribute significantly to HQOLI (P = 0.01), but chorea does not (P = 0.21). CONCLUSIONS: These data support the validity of the HQOLI in patients with HD. The ceiling effect limits the use of HQOLI in late stages but appears necessary to achieve good discrimination at early and intermediate stages. Correlations with other instruments suggest good external validity. High correlations of several items within dimensions suggest item number can be reduced. New, shorter versions of the questionnaire is in development. Assessments of responsiveness to change, test-retest reliability, and cross-cultural validation will be performed in parallel.