USE OF PRO INSTRUMENTS: COPYRIGHT ISSUES

Anfray C, Emery MP
Mapi Research Trust, Lyon, France

OBJECTIVES: To meet the continuing demand for PRO instruments in clinical research and practice more and more PRO instruments are being developed. Although the prerequisite for international pooling and comparison of data is an identifiable original instrument and an official translation in a given language, this is not always the case. The need to modify instruments and the emergence of item banks make it difficult to identify an original instrument and the access to it is complicated as international copyright law is not sufficiently clear on this point. The objective of this abstract is to review the issues encountered by our distribution centre when disseminating instruments and their translations in collaboration with their developers and to provide recommendations for those who wish to develop an instrument, access it, use and/or modify it.

METHODS: To establish recommendations we proceeded as follows: 1) review of all requests to access, use and/or modify an instrument or its translations; 2) classification of requests, and 3) recommendations. RESULTS: Out of the 2679 requests concerning PRO instruments in 2006, 6 types of questions emerge to what are the conditions to: 1) access; 2) translate; 3) reproduce; 4) use in e-application; 5) modify; 6) use selected items of a given instrument. In response to this international copyright law needs to be interpreted and solutions for conflicting laws are needed.

CONCLUSION: Despite its importance for clinical research the identification of and access to an original instrument and its translations is not easy. This is complicated by the emerging need to modify instruments and the absence of clear indications by international copyright law. Our findings indicate the importance of questionnaire distribution centres as a step in the direction of solving copyright issues.

HOW GOOD IS GOOD ENOUGH? INTERNAL VALIDITY OF STATED PREFERENCES FOR DRUG THERAPIES

Ozdemir S, Mohamed AF, Huiberg C, Johnson FR
Research Triangle Institute, Research Triangle Park, NC, USA

OBJECTIVES: To compare measures of internal validity across similar stated-choice surveys, evaluate consequences of inconsistencies, and determine whether there is a basis for establishing a minimum standard for validity. METHODS: Axioms of utility theory require that valid preferences be stable, monotonic, and transitive. Counts of validity failures were obtained for each survey respondent in 8 stated-choice or discrete-choice experiment surveys. Each survey was administered to adults who were diagnosed with a specific disease. Each survey required respondents to choose between 8 to 12 pairs of treatment alternatives with varying treatment features. We also identified respondents with non-compensatory or lexicographic preferences, where subjects choose alternatives based on a single attribute. Using negative-binomial regression we estimated the effect of personal characteristics on internal-validity test results. RESULTS: Choices from 3929 respondents were evaluated for internal validity. Across all surveys, approximately 20% failed at least one stability test, 10% failed at least one monotonicity test and 15% exhibited lexicographic preferences. Non-white respondents (p < 0.05) and respondents with less education (p < 0.05) were more likely to fail one of these tests. The effect of income was concave. As income increases the likelihood of failure of one or more tests decreases; however, above an annual income of $100,000, the likelihood of failure increases. Gender, age and employment had no significant effect on internal validity. With the exception of respondents with lexicographic preferences, removing respondents who fail internal validity tests generally did not materially change point estimates in preference models.

However, removing inconsistent respondents improved estimate precision by 1–5%. CONCLUSION: Our results suggest that a failure of non-lexicographic internal validity tests at a rate less than 25% does not bias preference estimates derived from stated-choice surveys. Identifying which respondent groups find stated-choice surveys challenging can help in developing and pretesting surveys that may contribute to widespread deployment of this questionnaire.

CROSS-OVER RANDOMIZED CONTROL TRIAL OF ELECTRONIC MEASUREMENT OF HEALTH-RELATED QUALITY OF LIFE IN CHINA

Chen T1, Li L1, Du Y1, Lei J2
Zhejiang University, Hangzhou, China, 1University of Goettingen, Goettingen, Germany

OBJECTIVES: Verification of the feasibility and reliability of the electronic version of Chinese SF-36 (based on the Quality-of-Life-Recorder) before its wide deployment. METHODS: Cross-over randomized controlled trial, comparing a paper based and an electronic version of the SF-36. According to generated random numbers, interviewees were asked to fill out either the electronic version or the paper version first. The second version was filled in after a pause of at least 30 minutes (medical students), or at least 10 minutes (patients). Convenience sample consisted of one group of 50 medical students and the other group of 100 patients. RESULTS: The acceptance of the electronic version was good (60% of medical students and 84% of patients preferred the electronic version). At the level of eight-scale scores, the mean-difference for each scale (except for General Health) between the two versions was less than 5%. At the level of 36 questions, the percentage of “Exact Agreement” ranged within 64–99%; the percentage of “Global Agreement” ranged within 72–99%; 77% of the Kappa coefficients demonstrated “good/excellent agreement” and 23% of the Kappa coefficients demonstrated “medium agreement”. CONCLUSION: Our findings support the feasibility and acceptability of an electronic version of the Chinese SF-36, as well as the agreement of results collected with paper-based and electronic version. The electronic versions may contribute to widespread deployment of this questionnaire.

FEASIBILITY, VALIDITY AND RELIABILITY OF THE WELSH VERSION OF THE EQ-5D HEALTH STATUS QUESTIONNAIRE

Hughes D
University of Wales, Bangor, Bangor, UK

OBJECTIVES: To investigate the feasibility, validity and reliability of the Welsh language version of the EQ-5D in a random adult population sample in Welsh-speaking areas of Wales. METHODS: 1000 names were selected at random from electoral registers, and questionnaires that included the Welsh version of the EQ-5D together with questions on socio-economic, demographic and health status, were mailed. Subjects were asked to rate their health on the day of completion. Respondents to the first questionnaire were sent a second EQ-5D within a fortnight to assess test-retest reliability. A reminder was sent if the second questionnaire was not returned within a fortnight. A pre-paid envelope was provided in each case. The acceptability and feasibility was assessed by examining the number of missing responses on returned forms. A reliability analysis was conducted, following the test-retest procedure, and construct valid-
ity was assessed in terms of convergent validity. RESULTS: 126 subjects returned the first questionnaire, of which 113 completed the second. Women represented 59.9% of the sample, and respondents’ mean age was 56.8 +/- 18 years. 38.7% of subjects reported having chronic illness. The number of missing items from the descriptive EQ-5D (aggregated for both administrations) was 4, representing 0.33% of all responses. Five EQ-VAS responses were returned incomplete. 97 (85.8%) respondents reported no change in health in the fortnight between the first and second administration. The Spearman rank coefficients for these respondents were 0.828 for EQ-5D scores, and 0.815 for EQ-VAS, respectively (both p < 0.000). The results of the regression analysis for convergent validity suggested that, together, EQ-VAS score, self-reported health (5-point Likert scale; poor to excellent), presence of chronic illness, and receipt of treatment, explained 66.0% of the variation in EQ-5D scores. CONCLUSION: It is concluded that the Welsh version of the EQ-5D has good acceptability, validity and reliability in measuring health status in subjects across Wales.

**OBJECTIVES:** To compare mapped utility instrument MID values for Hispanic respondents in the US national survey with a mean absolute difference of 8.5. A linear regression model was used to examine the relationship between utilities for the most prevalent self-assessed health in the two countries took the form EQ-5D_CHILE = 0.773*EQ-5D_US-H + 10.301 CONCLUSION: This research compares mapped utility instrument MID values within four newly identified patient groups. METHODS: Retrospective analysis of a two-year longitudinal study in a Western US managed care population was conducted. Individuals receiving a prescription for a new disease state in either year were selected. Four major disease states, each with more than 100 patients reporting a minimal change in health over the past year on the SF-36 general health question were included. MID values were calculated for gains and losses in health on the Brazier, Lundberg, Nichol, and Shmueli mapped utility instruments. MID values between disease groups for each instrument were tested with ANOVA. Effect sizes were compared in accordance to Norman (2001) as a distribution-based method to determine the MID. RESULTS: Results are displayed for Brazier’s SF-6D measure as all mapped utility instruments had similar results. 145, 240, 306, and 150 patients reported a minimal gain or loss in health over the year when they filled a new prescription for cardiac disease, COPD, depression and rheumatoid arthritis, respectively. The reported mean MID utility change for gains was 0.026, 0.022, 0.019, and 0.018 for each patient group, respectively (P = 0.975). The reported mean MID utility change for losses was -0.069, -0.045, -0.055, and -0.028 for each patient group, respectively (P = 0.217). The average effect sizes ranged from 0.133 and 0.194 for gains and from -0.241 and -0.627 for losses. Therefore, according to Cohen’s classification system, the MID values in gains were not significant changes, while those in losses were considered small. CONCLUSION: This data demonstrates consistent, but small, MID values across various disease groups for each utility instrument. Researchers should confirm results in other populations.

**OBJECTIVES:** It is unknown whether minimal important differences (MID) identified on generic patient reported outcomes are consistent across disease states. This research compares mapped utility instrument MID values within four newly identified patient groups. METHODS: Retrospective analysis of a two-year longitudinal study in a Western US managed care population was conducted. Individuals receiving a prescription for a new disease state in either year were selected. Four major disease states, each with more than 100 patients reporting a minimal change in health over the past year on the SF-36 general health question were included. MID values were calculated for gains and losses in health on the Brazier, Lundberg, Nichol, and Shmueli mapped utility instruments. MID values between disease groups for each instrument were tested with ANOVA. Effect sizes were compared in accordance to Norman (2001) as a distribution-based method to determine the MID. RESULTS: Results are displayed for Brazier’s SF-6D measure as all mapped utility instruments had similar results. 145, 240, 306, and 150 patients reported a minimal gain or loss in health over the year when they filled a new prescription for cardiac disease, COPD, depression and rheumatoid arthritis, respectively. The reported mean MID utility change for gains was 0.026, 0.022, 0.019, and 0.018 for each patient group, respectively (P = 0.975). The reported mean MID utility change for losses was -0.069, -0.045, -0.055, and -0.028 for each patient group, respectively (P = 0.217). The average effect sizes ranged from 0.133 and 0.194 for gains and from -0.241 and -0.627 for losses. Therefore, according to Cohen’s classification system, the MID values in gains were not significant changes, while those in losses were considered small. CONCLUSION: This data demonstrates consistent, but small, MID values across various disease groups for each utility instrument. Researchers should confirm results in other populations.