PMC26

## APPLICATION OF THE FRAMEWORK FOR EVALUATING COMPLEX INTERVENTIONS TO CLUSTER RANDOMIZED TRIALS FOR THE EVALUATION OF DISEASE MANAGEMENT PROGRAMS

Marchisio S[1], Panella M[2]
[1]University Politecnica delle Marche, Ancona, Italy, [2]Univerisity of Eastern Piedmont "A. Avogadro", Novara, NO, Italy

Trials of disease management programs pose several methodological challenges. Our objective is to assess the extent to which the various development steps of a cluster randomized trial to evaluate disease management are represented in the framework for the design and evaluation of complex interventions. The framework for evaluating complex interventions developed by Campbell and colleagues is composed by five phases: theoretical, identification of components of the intervention, definition of trial and intervention design, methodological issues for main trial, and promoting effective implementation. Using these phases the corresponding stages in the development of the cluster randomized trial to evaluate the effectiveness of disease management programs are identified and described. Synthesis of evidence needed to construct the program, survey and qualitative research used to define components of the program, a pilot study to assess the feasibility of delivering the care, methodological issues in the main trial including choice of design, allocation concealment, outcomes, sample size calculation and analysis are adequately represented using the stages of the framework for evaluating complex interventions. Even though is difficult to define precisely what exactly the "active ingredients" of a program of disease management and how they relate to each other, we think that the applied framework is a powerful resource for researchers planning a randomized clinical trial to evaluate the effectiveness of such programs.

**WITHDRAWN** PMC27

## RESEARCH ON METHODS & CONCEPTUAL PAPERS—Patient-Reported Outcomes Studies

PMC28

## RASCH PARTIAL CREDIT ANALYSIS OF THE SF-12V2 USING THE 2003 MEDICAL EXPENDITURE PANEL SURVEY (MEPS)

Gu NY, Doctor JN
University of Southern California, Los Angeles, CA, USA

This study assesses the Rasch measurement properties of the SF-12 version 2 (SF-12v2) physical and mental health (PH and MH) items in respondents with most prevalent chronic conditions. Medical Expenditures Panel Survey (MEPS) respondents' age ≥ 18 with complete SF-12v2s from 2003 were extracted (n = 19,906). Eleven subgroups were identified using the primary ICD-9-CM code for the top 10 chronic conditions (hypertension, diabetes, depression, back disorder, arthropathy, cholesterol, asthma, sinusitis, anxiety and joint disorder) and healthy persons (n = 8324). Respondents with perfect scores demonstrating ceiling (n = 303) and floor effects (n = 12) were removed to ensure uncertainty in the responses. Coding reflected that higher scores represent healthier respondents. The Rasch partial credit model was used to exam the item category properties and fit statistics. Residual factor analysis was conducted to assess the factor loadings on the items. Item misfit took place mostly among MH items (infit/outfit z-score >2.0). Particularly, MH item "Have you felt calm and peaceful?" showed misfit across all subgroups. Rasch residual analysis identified 75.2% to 86.9% of the shared variance within clusters of all items. Further, PH items had positive factor loadings while MH items had negative factor loadings, with few exceptions (3/11). Some item category steps had poor step properties. Mental health items are more likely to be noisy. The patterns of the factor loadings of the PH and MH items were as expected in most of the groups suggesting that the SF-12v2 has two distinct factors measuring the overall health. Poor item category performance suggests that collapsing of these categories might improve the quality of the instrument.

PMC29

## INTERNATIONAL VALUATION SET FOR EQ-5D HEALTH STATES

Craig BM[1], Busschbach JJ[2], Salomon J[3]
[1]Moffitt Cancer Center, Tampa, FL, USA, [2]Erasmus Medical Center, Rotterdam, Netherlands, [3]Harvard University, Boston, MA, USA

**OBJECTIVE:** Health states defined by classification systems like the EQ-5D can be valued using techniques that aim to elicit cardinal measures directly, such as the time trade-off (TTO) and visual analogue scale (VAS), but also by ranking the health states. In this study, we estimate international value sets for the 243 EQ-5D states based on rank, TTO and VAS responses and test their equivalence. **METHODS:** We estimated the coefficients scale is recommended in Sweden [in Swedish] of a conditional logit and a linear probability model of rank responses as well as the coefficients of two linear models of TTO and VAS responses using pooled data from eight countries: Slovenia, Argentina, Denmark, Japan, Netherlands, Spain, UK and US, which gave us 179,431 responses from 11,483 subjects. The main difference between the two rank models is that one models utility in ln(odds), as suggested by McFadden, and the second in probability. The regression specifications have previous models estimated in the United States and UK nested within its framework. Furthermore, we compare rescaled predicted values for the 242 EQ-5D states, excluding 11,111, across techniques in terms of correlation and concordance. **RESULTS:** The non-optimal gap reduces when utilities are linearly modeled in probability instead of log-odds, as suggested by McFadden. The rank-based values are highly correlated with both TTO and VAS values. Compared to the log-odds model, the linear probability model produces rank-based values with greater concordance to TTO/VAS values. **CONCLUSION:** In former investigations we tested if ranking in the large pool of data produce values of health states comparable with direct measures as TTO and VAS. In this investigation we provide further evidence by showing convergent validity between TTO/VAS and alternative rank-based models for the whole valuation space. This evidence emphasizes the promise of a valuation technique that incorporates ordinal responses. Overall, rank-based values merits closer investigation.

PMC30

## A REVIEW AND CRITIQUE OF METHODS FOR MEASURING TEMPORARY HEALTH STATES IN COST-UTILITY ANALYSES

Wright DR[1], Wittenberg E[2], Swan JS[3], Miksad R[4], Prosser L[5]
[1]Harvard School of Public Health, Boston, MA, USA, [2]Brandeis University, Waltham, MA, USA, [3]Massachusetts General Hospital, Boston, MA, USA, [4]Beth Israel Deaconess Medical Center, Boston, MA, USA, [5]Harvard Medical School, Boston, MA, USA

**OBJECTIVES:** Temporary health states (states with a duration of less than one year) are common and include many infectious diseases, short-term treatments, and diagnostic procedures. Valuation of these states requires special consideration because the health state is transitory and standard methods ignore the influence of duration on preferences. Inaccurate assessments could introduce bias into cost-utility ratios. There is no "gold stan-

dard" valuation method. The aim of this study is to review and critique temporary health state valuation methods and identify areas for future research. **METHODS:** We reviewed the literature and evaluated preference-based temporary health state valuation methods according to five criteria: (1) Consistency with quality-adjusted life year theory; (2) Ease of use; (3) Relevance to temporary health state-specific domains; (4) Sensitivity to health state duration; and (5) Extent of bias. Our goal was to provide a critical assessment of methods that could be used to obtain values for use in cost-utility analyses. **RESULTS:** We identified six temporary health state valuation methods. Methods modified standard approaches by prorating utilities, using a chained approach, or trading-off waiting time or sleep instead of death in a time trade-off. These modifications capture the effect of duration better than standard methods. The strength of methods varied. No method was well tested for validity and reliability with respect to temporary health states. **CONCLUSION:** The literature on temporary health state valuation methods is sparse and inadequate. Our critique did not identify a method that is appropriate for valuation of all temporary health states. Selection of the most appropriate method should depend on the duration of and type of temporary health state being considered. Further research should focus on the validity, reliability and feasibility of valuation under different circumstances. Utility values obtained using temporary health state methods should be compared to those using standard methods to quantify biases.

**PMC31**

### CONTROLLING MEASUREMENT ERROR OF PATIENT-REPORTED-OUTCOMES DURING THE IMPLEMENTATION STAGE OF CLINICAL TRIALS

Gnanasakthy A

Novartis Pharmaceuticals, East Hanover, NJ, USA

**OBJECTIVE:** Measurement errors may be introduced in the development, cultural adaptation, implementation, and analysis of PRO assessments. Recent publications provide guidance to minimize measurement errors during the development and cultural adaptation stages. Very little guidance is available to control errors, especially in multinational studies, introduced during the implementation of PRO assessments. The objective of this abstract is to highlight errors that may be introduced during the implementation stage, specifically during the production of data capture modules (e.g. Case Report Forms)for multi-national studies. **METHODS:** A rigorous process was put in place to monitor errors introduced during the CRF development process with the aim of having a library of PRO instruments readily available for use in clinical trials. After typesetting, CRF pages were proof read by three independent reviewers including a native speaker. Suspected errors including poor grammar and typographical errors found in original PRO instruments were reconciled with author's permission and documented. **RESULTS:** A total of 40 PRO instruments were used in 39 Phase III multinational studies involving 69 languages in 2006–2007. Three instruments had multiple versions for the same language and the author of another instrument did not have a list of available translations. Two types of errors were found at the final stage of proof reading by native speakers. The first, ambiguous or outdated terminology. The second, typesetting errors which may have altered the meaning of the phrase or question. **CONCLUSION:** An adequate process must be in place to monitor, document and minimize errors that may be introduced during the implementation stage of PRO assessments. Failure to do so, especially in multi-national studies, may invalidate the resources spent during the development and translation stages and increase the Company Risk.

**PMC32**

### PREDICTING SF-6D PREFERENCE-BASED UTILITIES USING MEAN SF-36 HEALTH DIMENSION SCORES WHEN PATIENT LEVEL DATA ARE NOT AVAILABLE

Ara R[1], Brazier JE[2]

[1]University of Sheffield, Sheffield, South Yorkshire, UK, [2]The University of Sheffield, Sheffield, South Yorkshire, UK

**OBJECTIVES:** The objective of the study is to derive an algorithm to predict a cohort preference-based SF-6D index using the eight mean health dimension scores when patient level data is not available. **METHODS:** Health related quality of life data (n = 6890) collected from patients with a wide range of health conditions was used to explore the relationship between the SF-6D and the eight dimension scores. Ordinary least square regressions were derived using the eight dimension scores and first order interactions. Models were assessed for goodness of fit and predictive abilities using standard statistics such as variance explained; residuals and the proportion of predicted values within the minimal important difference. The models were also compared on their abilities to predict mean cohort SF-6D scores using mean dimension scores using both within-sample and out-of sample published datasets. **RESULTS:** The OLS equations obtained explained over 83% of the variance in the individual SF-6D scores. While the models over-predict the lower health states and under-predict the higher SF-6D scores on the individual level, the mean absolute errors are in the region of 0.040. When using mean dimension scores from within-sample subgroups and out-of sample published datasets, the majority of predicted scores were well within the minimal important difference (0.041) for the SF-6D. The models are reasonably accurate at predicting incremental values between study arms (mean error 0.012; mean absolute error 0.017) and when predicting incremental changes over time (mean error 0.004; mean absolute error 0.024). **CONCLUSION:** This paper presents a mechanism to estimate a mean cohort preference-based SF-6D score from published mean dimension scores. This study is unique in that it uses published mean statistics to validate the results. The out-of sample validation demonstrates the algorithms can be used to inform both clinical and economic research. Further research is required in different health conditions.

**PMC33**

### PREDICTING A MEAN EQ-5D PREFERENCE-BASED SCORE FROM THE 8 MEAN SF-36 DIMENSION SCORES WHEN INDIVIDUAL DATA IS NOT AVAILABLE

Ara R[1], Brazier JE[2]

[1]University of Sheffield, Sheffield, South Yorkshire, UK, [2]The University of Sheffield, Sheffield, South Yorkshire, UK

**OBJECTIVES:** The objective of the study is to derive a method to predict a cohort EQ-5D preference-based index score using published statistics of the eight dimension scores describing the SF-36 health profile. **METHODS:** Ordinary least square regressions are used to obtain models from patient level data covering a wide range of health conditions. The eight dimension scores, the squares age and gender are used to derive a relationship with the EQ-5D index. Models obtained are compared for goodness of fit using standard techniques such as descriptive statistics, variance explained, the residuals and the proportion of values within the minimal important difference. Predictive abilities are also compared when using summary statistics from both within-sample subgroups and datasets published studies. **RESULTS:** The models obtained explain more than 56% of the variance in the EQ-5D scores. For the individual predicted values, the mean predicted EQ-5D score is correct to two decimal places and the mean absolute error is approximately 0.13. Using summary statistics to