Approach to Fixing the Discrepancy Found in the Wu et al. FACT-P to EQ5D Mapping Algorithm—Reply to Letter to the Editor by David Cella, Irina Proskorovsky, and Feng Pan

To the Editor – We appreciate Cella et al.’s letter and their interest in the algorithm for predicting EuroQol five-dimensional (EQ-5D) questionnaire scores using Functional Assessment of Cancer Therapy - Prostate (FACT-P) questionnaire scores, published in Wu et al. In the letter, the authors raised two primary issues with the algorithm and proposed a modification. We welcome the opportunity to reply to these concerns.

First, the authors pointed out that while using the proposed algorithm, the predicted mean EQ-5D values were out of range. In response to this issue, we would like to emphasize a crucial step in the algorithm. Table 4 on page 412 provides the final full algorithm with and without EORTC. The algorithm, however, also requires a truncation of the predicted value of EQ-5D at the final step to ensure the correct range of the predicted value. This has been mentioned in page 410 of the published article:

If the predicted value of EQ-5D fell outside the defined range of [-0.594, 1.000], then it was truncated to the appropriate boundary value.

Note that the proposed model is not a linear model but a truncated linear model in which the predicted values are truncated with a floor and ceiling (-0.594, 1.000). Using only the linear model with the coefficients reported in Table 4 on page 412 can produce mean EQ-5D values outside that range; hence, it is crucial to use the truncation step of the algorithm. We believe that using the prediction algorithm along with this truncation step will ensure mean EQ-5D values within the correct range.

Second, we acknowledge the typographical error made on page 410 (“an average BMI of 72.4 (SD = 9.0)” as pointed out by the authors. Based on the original result of this study, the correct summary statistics for BMI should be mean = 27.3 and SD = 4.1. This typographical error, however, affects only this particular sentence of the article and does not by any means affect the validity of the algorithm. We are really sorry for this mistake and any confusion this may have caused to the readers, including the authors, and to Value in Health.

Once again, we thank the authors as well as the editor for identifying the typographical error. After reviewing the original data, we believe that the published algorithm is correct. We sincerely hope we were able to address the concerns the authors raised and were able to emphasize the key step in using the published algorithm in a correct way to predict utility.

Source of financial support: The authors have no other financial relationships to disclose.

Eric Q. Wu, PhD, MS, MA
Analysis Group, Inc., Boston, MA, USA

Gourab De, PhD
Analysis Group, Inc., New York, NY, USA

1098-3015/$36.00 – see front matter
Copyright © 2012, International Society for Pharmacoeconomics and Outcomes Research (ISPOR).
Published by Elsevier Inc.
http://dx.doi.org/10.1016/j.jval.2012.03.1391

Improving the Measurement of QALYs in Dementia: Some Important Considerations

To the Editor – We welcome the efforts of Mulhern et al. [1] to improve the estimation of quality-adjusted life-years in dementia. Indeed, we have previously argued that instruments measuring dementia-specific health-status utilities would represent a major step forward in dementia research [2]. Nonetheless, we have some reservations about certain aspects of the DEMQOL-U and DEMQOL-Proxy-U developed by these authors. In particular, we have concerns about the content validity of the DEMQOL-U and the analytical strategies applied by its developers.

Content Validity
Our first concern is regarding the items Mulhern et al. have selected for their health-state classification system (DEMQOL-U). We question the content validity of the items, because these do not cover the full spectrum of dementia health-related quality of life (HRQOL). In our opinion, the authors place too much emphasis on mood-related items. It is generally accepted that health is composed of three domains: physical, mental, and social [3,4]. One would therefore expect any HRQOL measure to cover all three to at least some degree. Disease-specific HRQOL measures will most likely put more emphasis on one or two of the three domains depending on the disease. In the case of dementia, one would expect an instrument to emphasize mental and social well-being. This is exactly what the original DEMQOL does. In contrast, not all these domains are covered by the DEMQOL-U.

The authors of the original article describing the development of the DEMQOL measure used a conceptual framework of five do-
mains: 1) daily activities and looking after yourself, 2) health and well-being, 3) cognitive functioning, 4) social relationships, and 5) self-concept. By comparison, three of the five items of the DEMQOL-U fall under “health and well-being” and two under “cognitive functioning.” This suggests that domains 1, 4, and 5 that Mulhern et al. intended to cover were omitted from their classification system. In the article by Smith et al. [5], the authors present a preliminary factor analysis solution for the field test data. It covers four factors: 1) daily activities, 2) memory, 3) negative emotion, and 4) positive emotion. The final field test revealed a different four-factor structure, which the authors found more difficult to interpret.

The study by Mulhern et al. advances a five-factor solution replicating the original factors 2, 3, and 4. In addition, it identifies two new factors, Social relationships and Loneliness. The omission of a domain that describes physical functioning (and thus impacts daily activities), however, is a major concern. Dementia denotes a class of illnesses that occur mostly in frail elderly people. Accordingly, many patients with dementia suffer from physical comorbidity. Thus, the omission of a physical domain could lead to the overestimation of patient-reported utilities.

The absence of several relevant conceptual domains is not our only concern. The items covering cognition and relationships that are part of the DEMQOL-U classification system might indicate some aspects of “worrying” in addition to or instead of the intended item content because of the way they are framed. The DEMQOL measure was framed in aspects of worrying because this stem was most easily understood during pretesting. The authors of the DEMQOL, however, allocated “being worried or anxious” to the domain of health and well-being. In that light, framing separate items in terms of “being worried about . . .” raises the possibility of confounding for these items.

Analytical Strategies

Our second concern is regarding the analytical strategies applied by Mulhern et al., namely, factor analysis and Rasch analysis. Factor analysis seems unnecessary, because the same technique was used to develop the original DEMQOL instrument, albeit the number of observations was substantially lower in the earlier study. This may explain the differences between the solutions found in each study. It should be noted that factor analysis is fully directed by relationships (i.e., correlations) between variables (i.e., items). This means that two items with more or less equal distributions of responses (i.e., frequencies) will load on the same factor. However, factor analysis results will not tell us anything about the weight (i.e., importance) of these items.

Subsequently, the authors apply Rasch analysis on the items for each factor derived by the factor analysis. In the Rasch analysis, they perform several tests, one of them testing for the unidimensionality of each dimension. This does not yield much information, because the basic feature of factor analysis with varimax rotation is that it produces orthogonal (i.e., unidimensional) factors. Apart from this, the standard Rasch analysis may not be the correct response model for the type of data that are obtained in the setting of reported health levels. Mulhern et al. apply the Rasch model to Likert items, although these items do not have the correct response structure. We have noticed this incongruence in many other studies directed at transforming descriptive HRQOL questionnaires or instruments into preference-based HRQOL instruments [6–10]. Rasch analysis requires a “cumulative” data structure (if a respondent agrees with a statement of a certain level, this means that this person also agrees with the statements that precede this level). In standard descriptive HRQOL questionnaires, we are dealing with an “ideal point” or “single-peaked” data structure. When persons whose attitudes are to be measured agree or disagree with a statement, the implied response function is single-peaked. In other words, it is expected that a person will agree with the statements that are close to the person’s own attitude and disagree with those statements (e.g., categories of the item) that are far from the person’s location on the scale in either direction.

Coombs [11] developed this implied response process within a deterministic framework and coined the term “unfolding” for the simultaneous processes of locating persons and items on a scale from the agree/disagree responses. This term continues to be used in the literature. Unfolding, however, became extremely cumbersome for more than four statements. Therefore, it did not pose a challenge to the Likert approach as the favored procedure in practice. Nonetheless, modern extensions of this unfolding model that can deal with a large number of statements now exist. Given the response options for the items in the DEMQOL, we believe that such a probabilistic unfolding model for polytomous responses may have been a more valid method for item selection [12].

In summary, we feel that the introduction of the DEMQOL-U represents an important step in the right direction. The instrument, however, still has two important weaknesses: insufficient comprehensiveness and limited validity of data analysis. These weaknesses require further consideration before its use in research and clinical practice is warranted.

REFERENCES

Response to Comments on Mulhern et al., “Improving the Measurement of QALYs in Dementia: Developing Patient- and Carer-Reported Health State Classification Systems Using Rasch Analysis”

To the Editor – We thank Arons et al. [1] for their comments regarding our article using traditional psychometric and Rasch analysis to develop dementia-specific health-state classification systems (DEMQOL-U and DEMQOL-Proxy-U) [2] that have subsequently been valued to generate utility scales for use in the generation of quality-adjusted life-years (QALYs) [3]. We take this opportunity to respond to their reservations regarding the content validity of DEMQOL-U, and the analytical techniques we used to develop the classification systems.

Content Validity

Arons et al. criticize the content validity of the domain structure used in DEMQOL-U, in particular the omission of a physical domain (that impacts daily activities).

Although the conceptual framework that underlies DEMQOL and DEMQOL-Proxy included five domains (daily activities and looking after yourself, health and well-being, cognitive functioning, social relationships, and self-concept), these five domains were not fully supported by the factor analysis carried out for both the preliminary field test and the final field test during the original development process [4,5]. The final (item-reduced) version of DEMQOL did not fully represent daily activities (two items) or self-concept (0 items). We did not therefore intend to cover the original five conceptual domains in DEMQOL-U.

We instead carried out exploratory factor analysis by using a much larger data set (n = 644) and established a different five-factor structure for DEMQOL including 1) cognitive functioning, 2) positive emotion, 3) negative emotion, 4) social relationships, and 5) loneliness. We would argue that using a significantly larger data set in a population of patients with a definite diagnosis of dementia provides a stronger empirically based representation of the underlying factor structure of DEMQOL than was established by earlier analysis. Furthermore, in all the factor analysis carried out on DEMQOL [2,4,5], the well-being items have consistently split into positive and negative emotion, proving support for including them as separate dimensions of DEMQOL-U. Factors 1 to 3 assess mental well-being, and factors 4 and 5 assess social well-being, which, as suggested by Arons et al., are key domains to include in any measure of dementia-related health-related quality of life (HRQOL).

Qualitative work carried out by Smith et al. [4, 5] suggested that the daily activities items were more important when reported by a proxy than by people with dementia themselves. Arons et al. have focused on DEMQOL-U, but the DEMQOL-Proxy-U classification system includes an appearance dimension, which is related to daily activities. The difference in the two classification systems serves to highlight one of the reasons why both DEMQOL and DEMQOL-Proxy should be used as complementary measures.

Although the physical aspect of HRQOL is likely to be an important area for many patients with dementia, Arons et al. point out that a dementia measure should emphasize mental and social components of HRQOL. A condition-specific measure should focus on those dimensions most related to the condition and the outcomes of its treatment. We argue that this is exactly what DEMQOL-U does. Therefore, the utility values generated by DEMQOL-U are valid for assessing key outcomes of any dementia-related intervention or treatment. It is also expected that the system will be used alongside other instruments that would assess co-occurring aspects of HRQOL. Problems with physical aspects of HRQOL arising from comorbidities will not impact on the size of the change in health-state value as a result of an intervention provided that there are no preference interactions with the mental and cognitive dimensions. Recent evidence from the valuation of a mental health measure (Clinical Outcomes in Routine Evaluation [CORE-6D]) supports the notion that the relationship between physical and mental health is additive [6,7].

Arons et al. also criticize the cognition and social relationships items included in DEMQOL-U because of the question stem used in the original instrument, which frames the question in terms of worrying about the concept (e.g. “How worried have you been about forgetting things that happened recently”). DEMQOL includes an item directly asking about worry (“Have you felt worried”) that in the revalidation factor analysis fitted with the negative emotion factor. Because of the focus on worry, Arons et al. suggest the possibility of confounding. In the development of DEMQOL, framing the questions in terms of worry about a particular concept was the stem that was most easily understood and so it was used for 15 of the 28 items. The original DEMQOL measures have been shown to be psychometrically valid and reliable by using this question stem, and therefore provide a valid basis for the development of a utility measure based on the standard question wording.

Analytical Strategies

Arons et al. also criticize the analytical strategies we used, which involves factor and Rasch analysis as part of a six-step process.