Predicting EQ-5D responses from SF-12: should we take into account dependence and ordering?

Caterina Conigliani, Andrea Tancredi, and Andrea Manca

Abstract Generic health status measures such as the SF-12 provide important informations about health-related quality-of-life (HRQL), but they do not incorporate preferences for health states and cannot be used for the calculation of quality-adjusted life-years (QALYs). It follows that in order to conduct a cost-effectiveness analysis it is common to use mapping algorithms to estimate preference-based HRQL intruments scores from SF-12 scores. Here we consider the problem of directly predicting EQ-5D responses rather than utility values, and in recognising that there might be dependence between the five dimensions of EQ-5D responses, and that the possible levels of each dimension are ordered, we explore the behaviour of a multivariate ordered probit regression model.

Key words: health status, SF-12, EQ-5D, multinomial logit, multivariate ordered probit

1 Introduction

Health economists are often interested in deriving health state utilities from disease-specific health status measures. This problem includes both the direct elicitation of utility values from responses to health-related quality-of-life (HRQL) instruments (see, for instance, [3]) and the mapping between different instruments where one of them has an existing set of utility values (see, for instance, [5]). In particular, most of

C. Conigliani
Dipartimento di Economia, Università Roma Tre, e-mail: caterina.conigliani@uniroma3.it

A. Tancredi
Dipartimento di Metodi e modelli per l'Economia, il Territorio e la Finanza, Università di Roma La Sapienza, e-mail: andrea.tancredi@uniroma1.it

A. Manca
Centre for Health Economics, University of York, e-mail: andrea.manca@york.ac.uk

the studies that have considered this latter approach have used ordinary least squares regression, that as pointed out for instance in [7] in this setting has several potential limitations. In fact, it implies that utilities are continuous random variables, so that the probability of full health is small, and the presence of ceiling effects leads to inconsistent estimates of the regression coefficients. For these reasons, [7] consider a different approach: for the particular problem of mapping between SF-12 and EQ-5D, they suggest directly predicting responses rather than EQ-5D utility values.

Recall that the EQ-5D is a standardised, non-disease-specific instrument developed by the EuroQol Group (see, for instance, [2]) for describing and valuing health-related quality of life. It has been developed for self-completion, and it requires respondents to describe their own health state on five dimensions regarding mobility, self-care, usual activity, pain/discomfort, anxiety and depression, each of which has 3 levels (no problem, some problems, extreme problems), so that each health state can be identified by a five-digit number. Moreover, the EQ-5D instrument has been purposefully designed for generating a cardinal index of health, thus giving it substancial potential for use in health care evaluation. This is done by asking the respondent to mark off his current health state on a visual analogue rating scale (VAS), i.e. a vertical scale where the endpoints are labelled *Best imaginable health state* and *Worst imaginable health state*; utility value sets have then been obtained for all possible 243 health states using the EQ-5D VAS technique ([2]) and the time trade-off method ([4]).

The main point is that the EQ-5D valuation questionnaire is mainly distributed in instances where researchers specifically wish to elicit valuations of health. Other instruments, such as the SF-36 (which is a multi-purpose health survey with 36 questions) or the SF-12 (that contains a subset of items from the SF-36), do not incorporate preferences for health states, and cannot directly be used for the calculation of QALYs and for health care evaluations. This motivates the need for mapping algorithms that allows to estimate preference-based HRQL intruments scores from generic health status measures.

Going back to the particular problem of mapping between SF-12 and EQ-5D, the *response mapping* approach of [7] employs multinomial logit regressions to estimate the probability that a respondent selects a particular response level for each question, then uses Monte Carlo simulation to allocate respondents on one of the EQ-5D discrete health states, and finally calculates utility values from this set of predicted responses, in this way preserving the main design features of the EQ-5D instrument. Here we consider a similar approach: for a large population survey in which both the SF-12 and the EQ-5D were administered, we look at the prediction of EQ-5D responses, and then compute the corresponding utility values and compare them with the actual EQ-5D results reported in the survey. As in [7], as explanatory variables we consider the physical (PCS-12) and mental (MCS-12) summary scores derived from SF-12, together with their squares and interactions. Aim of the analysis is to explore the use of models that, unlike that suggested in [7], take into account both the dependence between the dimensions of EQ-5D responses and the ordering among the possible levels of each dimension. The whole analysis is carried within the Bayesian framework employing Markov Chain Monte Carlo tecniques.

2 The models

As we pointed out earlier, in the response-mapping approach of [7], multinomial logistic regression was used to explore the association between responses to the SF-12 and responses to the EQ-5D.

Let y_{ij} be the EQ-5D response of individual i to dimension j ($i = 1,...,n$, $j = 1,...,5$), and let $y_i = (y_{i1},...,y_{i5})$ be the vector of responses of individual i. Moreover, let x_i be the vector of explanatory variables observed on individual i and β_{lj} be the corresponding vector of coefficients for level l ($l = 1,2,3$). Then the probability of observing outcome l as a function of the linear combination $x_i\beta_{lj}$ can be written as

$$Pr(y_{ij} = l|x_i) = exp(x_i\beta_{lj}) / \sum_{s=1}^{3} exp(x_i\beta_{sj}),$$

where for identification purposes we set $\beta_{1j} = 0$ ($j = 1,...,5$), so that the remaining logit coefficients represent change relative to the $y_{ij} = 1$ outcome.

However, it is emphasised for instance in [1] that many advantages can be gained from treating an ordered categorical variable as ordinal rather than nominal, and the key point is that ordinal variables are inherently quantitative; it follows that is often more crucial to conclusions the distinction regarding whether data are nominal or ordinal than the choice between a model that recognise the discrete nature of categorical data, such as the multinomial distribution, and a continuous sampling model. This is the first motivation that lead us to take into account the ordering among the possible levels of each dimension when considering the problem of mapping between SF-12 and EQ-5D responses. The second motivation is related to the possibility to allow some form of dependence between the different EQ-5D dimensions. In fact, if we do not give some *structure* to the multinomial baseline model, then it would be extremely difficult to take into account dependence, since the number of parameters that need to be estimated would be enormous. However if we consider the ordering among the possible levels of each dimension, then we can take into account also the dependence more naturally.

In what follows we will compare the multinomial logistic regression of [7] with two alternative models. The first one assumes an independent ordered probit model for each of the 5 EQ-5D dimensions, and can be easily written in terms on an ordinary regression model for an underlying latent variable: $z_{ij} = x_i\beta_j + \varepsilon_{ij}, \varepsilon_{ij} \sim N(0,1)$ where $y_{ij} = l$ if $\gamma_{j,l-1} < z_{ij} \leq \gamma_{j,l}$ ($i = 1,...,n$; $j = 1,...,5$; $l = 1,2,3$) and where, for identification purposes, we set $\gamma_{j,0} = -\infty$, $\gamma_{j,1} = 0$, and $\gamma_{j,3} = \infty$.

The second alternative model is a multivariate ordered probit, that takes into account both the ordering and the dependence, and that can also be written in terms of latent variables by assuming that $\varepsilon_i = (\varepsilon_{i1},...,\varepsilon_{i5}) \sim MN(0,R)$. Here the unknown parameters are the β_j, the $\gamma_{j,2}$ ($j = 1,...,5$) and the elements of correlation matrix R. Posterior simulation for this model can be done, for example, by following the *parameter expanded- data augmentation* algorithm proposed by [8].

3 Results

The main purpose of the National Health Measurement Study (NHMS) was to compare commonly used HRQL instruments when they were co-administred to a cross-sectional sample of U.S. adults [6]. In particular, the NHMS collected data on 3844 adults from four different questionnaires, together with various other demographic and socioeconomic variables that were thought to be associated with HRQL.

The results of the comparison of the different models for the NHMS data are shown in Table 1, and show that the multivariate ordered probit is the preferred model. The comparison of the different models is based on both likelihood-based criteria of goodness of fit, such as the *Deviance Information Criterion*, and on the performance in terms of prediction; this is assessed by splitting the data set into two subsets that are used respectively for estimation and for validation, and by looking at both the mean square error (MSE) between observed and predicted EQ-5D utilities and the proportion of correct predicted responses on all EQ-5D dimensions (PCPR).

Table 1 The NHMS study: model estimation based on $n_1 = 2000$ observations, validation based on $n_2 = 1795$ observations

Model	**DIC**	**MSE**	**PCPR**
Independent univariate multinomial logit	8726	0.0137	0.2667
Independent univariate ordered probit	8808	0.0138	0.2635
Multivariate ordered probit	8527	0.0131	0.2800

References

1. Agresti A. Analysis of Ordinal Categorical Data, Wiley, 2nd Edition (2010)
2. Brooks, R, with the EuroQol Group. EuroQol: the current state of play. Health Policy **37**, 53-72 (1996)
3. Brazier J, Roberts J, Deverill M. The estimation of a preference-based measure of health from the SF-36. J Health Econ **21**, 271-92 (2002)
4. Dolan P, Gudex C, Kind P, Williams A. The time trade-off method: results from a general population study. Health Econ **5**, 141-54 (1996)
5. Franks P, Lubetkin EI, Gold, MR, Tancredi DJ. Mapping the SF-12 to preference-based instruments: convergent validity in a low-income, minority population. Med Care **41**, 1277-83 (2003)
6. Fryback DG, Dunham NC, Palta M, Hanmer J, Buechner J, Cherepanov D, Herrington S, Hays RD, Kaplan RM, Ganiats TG, Feeny D, Kind P. U.S. norms for six generic health-related quality-of-life indexes from the National Health Measurement Study. Med Care **45**, 1162-1170 (2007)
7. Gray AM, Rivero-Arias O, Clarke PM. Estimating the Association between SF-12 Responses and EQ-5D Utility Values by Response Mapping. Med Decis Making **26**, 18-29 (2006)
8. Lawrence E, Bingham D, Liu C, Nair VN. Bayesian inference for multivariate ordinal data using parameter expansion. Technometrics **50**, 182-191 (2008)