Universidade Vigo

Departamento de Economía Aplicada

http://webs.uvigo.es/x06

Working Paper: 13/01. November 2013

Development of a new preference-based instrument to measure dependency

Eva Rodríguez Míguez

José María Abellán Perpiñán

José Carlos Álvarez Villamarín

José Manuel González Martínez

Antonio Rodríguez Sampayo

Coordinator: Eva Rodríguez Míguez
emiguez@uvigo.es

Development of a new preference-based instrument to measure dependency*

Eva Rodríguez Míguez[†1]
José María Abellán Perpiñan[2]
José Carlos Álvarez Villamarín[1]
José Manuel González Martínez[1]
Antonio Rodríguez Sampayo[3]

Abstract

This paper reports the estimation of a preference-based scoring algorithm for a new dependency health state classification system. According to this system health states are described as a combination of 6 attributes (eat, incontinence, personal care, mobility, housework and cognition/mental problems), with 3 or 4 levels each. The tariff of this instrument is based on community preferences, hence it is consistent with the so-called 'societal perspective'. Preference weights can be used in QALY calculations and cost-utility analysis.

Keywords: Dependency, preference-based measures, QALY, time trade-off, Spain.

[†]Corresponding author: Universidad de Vigo. Departamento de Economía Aplicada. As Lagoas-Marcosende s/n. 36310 Vigo. Spain. E-mail address: emiguez@uvigo.es.
[1]Universidad de Vigo
[2]Universidad de Murcia
[3]Universidad de Santiago

*Financial support from Fundación Caixa Galicia, Consellería de Economía e Industria (Xunta de Galicia) grant no. 10SEC300038PR and Spanish Ministerio de Ciencia e Innovación grant no. ECO2011-25661 is gratefully acknowledged.

1. Introduction

Cost-utility analysis (CUA) provides a method for the economic evaluation of health care interventions. Health effects and cost increments are compared for alternative treatments, technologies or programmes. Health effects are assessed by calculating the number of quality-adjusted life-years (QALYs) gained with the intervention. In QALYs calculations life years are adjusted by preference weights (also called utilities) for health states.

Health state utilities can be obtained in two different ways. One possibility is to elicit preferences from the group of patients directly concerned with the intervention by using some measurement technique like, for example, the standard gamble or the time trade-off. This approach provides preference weights, but as it is usually recognized (Drummond et al., 2005) "is a very consuming and complex task". An alternative to the direct approach is to bypass the elicitation task by using one of the existing generic preference-based instruments. These instruments attach "off-the-shelf" utilities (preference weights elicited from a sample of the general population) to the descriptive health status information obtained from the patients. Well-known examples are EQ-5D (Dolan, 1997), SF-6D (Brazier et al., 2002, 2004) and Health Utility Index (Torrance et al., 1982, 1996; Feeney et al., 2002). As a result, these instruments are consistent with the so-called "societal perspective" (Gold et al., 1996), since they are based on community or general public preferences. This is a relevant feature as long as CUA is used to guide the allocation of resources on a societal level.

However, a disadvantage of generic preference-based measures is that their health state descriptive system is not sensitive enough for some medical conditions (Brazier et al., 1999), in such a way that effectiveness of interventions may be undervalued. This is the case for different chronic conditions (e.g. arthritis, diabetes, the effects of stroke, Alzheimer's disease) strongly associated with dependency, that is to say, with the inability to perform activities of daily living (ADLs) without the help of another person. Indeed, Donaldson et al. (1988) and Chilsholm et al. (1997) concluded that generic measures are, compared with condition specific measures, less sensitive to changes in older adults' health status. Hence, to optimally evaluate those health care programmes addressed to prolong or enhance independence amongst older adults a preference-based dependence-related utility measure would be required. In this article we present a measure developed with this aim.

It should be stressed that there are numerous different dependency scales (Kane and Kane, 2002) very useful for clinical purposes, but useless for CUA since they are not preference-based and, in consequence, are not able to yield QALYs. Given this limitation, some researchers (Sims et al., 2005; Bravata et al., 2005; Sims et al., 2008) have elicited preference weights for various ADL dependence health states based on the combination of the 6 ADLs included in the

Katz ADL (Katz et al., 1963) plus an additional ADL of walking (Goldstein et al., 2002). However, these authors did not estimate a scoring algorithm able to predict all the possible health states for dependence in ADLs. Moreover, there was no gradation of the interviewers' needs, but just participants classified themselves as either needing or not needing help to perform each ADL. Lastly, preferences were elicited from a convenience sample (older adult members of the Kaiser Permanente Medical Care Program of Northern California) not from the general population. Therefore, although the aforementioned studies are useful contributions in order to show that CUA of programmes that prevent or treat functional dependency should apply preference weights rather than relying on simple ADL counts, no "off-the-shelf" preference-based instrument was estimated to that end. We try to overcome that shortcoming by, firstly, constructing a new dependency health state classification system called DEP-6D and, next, estimating a scoring algorithm able to attach a preference weight to each of the all possible DEP-6D health states.

A similar approach to that we present in this article was previously followed by Ryan et al. (2006), though they were concerned with the evaluation of social care for older people, not with health care programmes. In consequence, the domains and levels of their instrument (OPUS) were designed to reflect whether needs relevant to social services clients are meet (e.g. whether their home is clean and comfortable) rather than to capture changes in functional status (e.g. whether dependency lowers with rehabilitation) as, indeed, is our aim. Another important difference between both works is that their utilities are not anchored at death, so their instrument is not able to yield 'generic' QALYs (i.e. QALYs comparable among different types of interventions) whereas ours can do it. Finally, their sample was opportunistic (people over the age of 60) rather than representative of the general population.

Other studies made in the realm of the health care and social services interventions for older people are far from our approach. In this sense, the ICECAP capability index for older people developed by Coast et al. (2008) does not assess preferences but, as its name suggests, value capabilities. Those values cannot be compared to standard QALYs produced by other interventions. In a turn, Burge et al. (2010) expand the attributes of the OPUS measure and obtain willingness-to-accept estimates for them. Thus, they do not estimate a utility algorithm but monetary estimates for different domains of social care output.

The paper is organized as follows. First, the construction of the descriptive component of the DEP-6D is described, as well as the valuation study and the estimation methods used. Next, main results are shown. A specific utility model for DEP-6D health states is recommended for economic evaluation purposes. A discussion closes the paper.

2. Methods

2.1. The DEP-6D dependency health states classification system

The process to design the descriptive component of the DEP-6D instrument was as follows. Firstly, there was an initial selection of attributes and impairment or need levels based on the review of the different dependency scales included in the Spanish National Health Survey. Next, those dimensions showing a high correlation amongst them were grouped into common categories in order to lower the number of potential attributes to be included in the classification system. Finally, a series of in-depth interviews with five experts in evaluation of ADL dependencies for the regional government of Galicia (north-western Spanish region) were conducted. These interviews allowed us to make the final selection of dimensions and levels as well as to identify potential unrealistic combinations among them. The resulting selection is showed in Table 1.

Table 1: The DEP-6D dependency health states classification system

Eat (EAT)	1. Does not need assistance to eat or drink. 2. Needs partial aid to eat or drink (cutting, serving, etc.). 3. Needs to be given food and drink.
Incontinence (INC)	1. Does not have incontinence or does not need help. 2. Has urinary incontinence (not fecal) and needs help for hygiene. 3. Has both urinary and fecal incontinence and needs help for hygiene.
Personal care (PER)	1. Does not need help for personal care: bathing, dressing, etc. 2. Needs help only to bath but not for the rest of his/her personal care. 3. Needs help for most personal care activities. 4. Is incapable of carrying out personal care. Needs someone to substitute him/her in this activity.
Mobility (MOB)	1. Moves independently. 2. Does not need help to move within the home but does out of home. 3. Needs help to move both in and out of home. 4. Is incapable of changing position. Bed-ridden or chair-ridden.
Housework (HOU)	1. Does not need help to carry out housework (cleaning, food, etc.). 2. Needs daily help for housework. 3. Is incapable of carrying out most tasks at home.
Cognition/mental problems (MEN)	1. Does not need help due to cognitive/mental problems or has not these problems. 2. Needs assistance to manage money, medication or to take some basic everyday decisions. Collaborative attitude with the care-taker. 3. Incapable of taking basic decisions. Cannot live alone. Does not offer resistance to help. 4. Incapable of taking basic decisions. Cannot live alone. Does not collaborate and usually offers resistance to help.

DEP-6D states are described as a combination of 6 attributes (eat, incontinence, personal care, mobility, housework and cognition/mental problems), with 3 or 4 levels each, so overall the classification system is able to yield up to 1,728 combinations in theory. A remarkable feature of the DEP-6D system is that includes cognitive/mental impairment as one of the attributes that characterizes dependency functional status. Cognitive impairment is not usually included in dependency scales although is a critical dimension to assess dependency in a global sense.

Think of people able to make most ADLs but that they do not act by their own initiative and do not collaborate with their caregivers. According to the interviewed experts, such situations would not be described in a suitable way just by means of the top 5 dimensions shown in Table 1; the sixth attribute (cognitive/mental problems) is added to capture this sort of dependency.

2.2. Selection of health states to be valued

The estimation of a scoring algorithm for the DEP-6D requires previously the direct valuation of a set of dependency health states by a sample of the general population. It is common in studies to model generic preference-based measures that the selection of the health states to be valued is based on an orthogonal design. However, the interviews conducted with experts revealed that some combinations of DEP-6D dimensions and levels may be implausible. In consequence, we opted by applying an optimal design (Fedorov, 1972) in order to exclude unrealistic combinations from the set of dependency health states to be valued ensuring, at the same time, the ability of obtaining accurate estimates for the remaining DEP-6D states. The OPTEX Procedure from SAS Software (version 9.1) was used to generate a set of 24 combinations (cards) divided into four blocks of size six. The D-efficiency of the design obtained was 75.5%. This design only captures main effects, so the existence of first and higher-order interaction effects cannot be tested. Notwithstanding, it has to be noticed that those models that allow for interactions between different dimensions do not usually improve the predictive ability of main effects models and frequently yield inconsistent estimates (Dolan, 1997; Greiner et al., 2005). Table 2 shows the states valued in each block.

Table 2: Dependency states (cards) evaluated by block*

Block 1	122222	Block 3	111112
	133334		113233
	211121		213322
	214232		222131
	313331		234431
	323433		334234
Block 2	111221	Block 4	123121
	112132		212223
	112211		233432
	223234		314434
	234333		324332
	333122		333231

* For a full description of each state see table 1. The number indicates the level of each attribute following the order of the table 1.

2.3. Selection of respondents to be interviewed

The sample consisted of 312 citizens drawn from the Galician general population recruited using a stratified random sampling. Face-to-face interviews were conducted by 6 trained interviewers. Each participant valuated only one of the four blocks of DEP-6D states shown in Table 2. Blocks were randomly assigned among all the participants. The order in which the six cards within each block were presented to each participant was randomized to minimize order effects. The average time per interview was around 20 minutes.

2.4. The questionnaire

The questionnaire was structured into five sections. Firstly, the DEP-6D classification system was explained to the respondents. Next, six dependency states were valued by means of the time trade-off (TTO) method (Torrance et al., 1972; Torrance, 1986). In the third part of the questionnaire, participants were asked to rank the six states previously valued. In the fourth section, respondents chose the first and second attribute they regarded as most severe among the six dimensions of the DEP-6D classification system. Lastly, standard sociodemographic questions (age, sex, income, education level, etc.) were asked to the participants.

2.5. The valuation method

As noted above the TTO technique was used to elicit individual preferences. Elicitation procedure began with a starting question to identify whether the state to be valued was regarded as better or as worse than death. Participants were asked to assume that they were seriously ill, in such a way that they would die unless they got a treatment. They had to make a choice between dying in a few days (No treatment) and spending 10 years in the dependency state being valued followed by death (Treatment). If they chose the treatment this meant that the state was regarded as better than death. On the contrary, if they refused the treatment then the state was regarded as worse than death. Depending on how the dependency health state was considered, the framing used for the TTO assessment was different.

If the state was regarded as better than death, the framing for the TTO consisted of the comparison between living 10 years in the dependency state (No treatment) and living X_{BDS} years in full health (Treatment). Next, an iterative up-down procedure based on standard decision analysis for 'zoomed in' on the indifference values in preference assessment (e.g. Keeney and Raiffa, 1993) was applied to find the number of years X_{BDS} at which the respondents were indifferent between the two options. The starting value for X_{BDS} was set equal to 5 years. Given that duration, respondents were asked to state if they would choose the treatment, would refuse it or if they would be indifferent between both options. In case they were indifferent between the two alternatives, the indifference point X^{*}_{BDS} was directly computed as the number of years in full health provided by the treatment option. On the contrary, if they chose one of the alternatives, then X_{BDS} was adjusted up or down until the indifference point was bounded. At that stage, the X^{*}_{BDS} was computed as the midpoint between the last answer given in a direction (e.g. No) and the previous answer given in the opposite direction (e.g. Yes). Imagine, for example, those respondents that do not choose the treatment when it offers 4 years in full health but that they do when it provides 5 years in full health. For these respondents X^{*}_{BDS} is taken as 4.5 years in full health. In case the participants regarded the dependency state as worse than death, the choice was between dying and spending X_{WDS} years in full health followed by (10 - X_{WDS}) years in the state being valued. As before, the value for X_{WDS} was initially set in 5 years and moved up or down until the convergence process terminated. Then the participant was introduced to the next choice scenario. In case participants were indifferent between the two alternatives, the indifference level was taken as the number of years in full health X^{*}_{WDS} provided by the treatment, followed by (10 - X^{*}_{WDS}).

Utility for state i, from participant j, y_{ij}, was obtained by assuming that the utility of each alternative can be decomposed according to the QALY model for chronic health states, i.e. $H(Q)\cdot T$, where $H(Q)$ is the utility function over health status and T are life years. In addition, the convention that the utility of full health is 1 and the utility of death is 0 was also adopted. If state i is regarded as better than death then, under the previous assumptions, $y_{ij} = X^{*}_{BDS} /10$, whereas if is regarded as worse than death, $y_{ij} = -X^{*}_{SWD} /(10- X^{*}_{SWD})$. However, since negative utilities calculated in this way do not have a lower bound, resulting in distributions very skewed to the left, we applied the transformation suggested by Patrick et al (1994), $y_{ij} = - X^{*}_{WDS} /10$, bounding negative values at -1.

2.6. Statistical inference

We estimated a main effects model by using random effects (RE) estimators. The general equation was defined as:

$$y_{ij} = \alpha + \sum_{d\epsilon D}\sum_{l\epsilon L}\beta_{ld}X_{ld} + \varepsilon_{ij}, \qquad [1]$$

where y_{ij} denotes the value that respondent i assigns to dependency state j; α is the intercept; X_{ld} represents the fifteen dummy variables, which indicate the presence of either level 2 or 3 or 4 (denoted as l) in a given dimension (d) in the state j; β are the parameters to be estimated; and ε_{ij} is the error term .

In the RE model the error term ε_{ij} is subdivided, such that:

$$\varepsilon_{ij} = u_j + e_{ij}, \qquad [2]$$

where e_{ij} is the unobservable error term due to differences among observations and u_i is the error term due to differences among respondents.

We also estimated extended versions of equation (1) including dummy variables denoting the presence in the state of the highest (lowest) level in at least one dimension, in a similar way to MOST (LEAST) terms for the SF-6D. The optimal specification was chosen according to the usual criteria of consistency, goodness of fit, and parsimony.

2.7. Analysis of ordering errors and invariant responses

Error/objection responses can be broadly classified (Witternberg and Prosser, 2011) into ordering errors (which include illogical and inconsistent responses) and objections/invariance (which include refusals to trade time or risk in preference elicitation questions). One source of ordering errors comes from logical, order or primary inconsistencies (Dolan and Kind, 1996; Devlin et al., 2003; Bravata et al., 2005). These *internal* inconsistencies arise when a logically worse health state is valued higher than a logically better health state (Badía et al., 1999). Another pattern of ordering errors occurs when rankings of health states predicted by valuations differ from those directly stated by the respondents. These *external* or criterion inconsistencies (Badía et al., 1999) are failures of convergent, empirical or external validity (Brazier and Deverill, 1999; Olsen et al., 2005; Abellán-Perpiñán et al., 2009), which means that utility scores elicited do not actually reflect preference orderings. In turn, invariance (Bravata et al., 2005) means that all the health states assessed by a respondent are given the same value. Refusals of respondents to tradeoff any lifetime or risk of death for quality improvement constitute a particular form of invariance, resulting in a utility score equal to 1. Invariance in utility assessments resembles embeddings effects in contingent valuation studies (Beattie et al., 1998).

Ordering errors and invariance were tested as follows. First, we explored the order consistency of TTO values by means of dominance tests (Ryan et al., 2006). We assume that an individual violates the ordinal dominance relationship between two dependency states (fails a dominance test) if the less severe state (dominant scenario) is valued lower than the more severe state (dominated scenario). Our design allowed us to identify pairs of dependency health states between which there was a dominance relationship given that DEP-6D attribute levels can be logically ordered. For example, as shown in Table 2, within block 1, state 211121 clearly dominates state 313331, because of the levels in all dimensions of the former are equal or lower (i.e. the severity of each of the dimensions is the same or better) than those of the latter. Overall, as shown in Table 2, there are 6 dominance comparisons into each block 1-3, and 4 in the forth block.

We also tested the external validity of TTO valuations by comparing the explicit ranking of the 6 cards directly ordered by each participant with the implicit ranking derived from the TTO values elicited from the same respondent. To measure the rank correspondence between responses, we estimated the Spearman’s rank correlation coefficient (or Spearman’s rho) between explicit and implicit rankings for each respondent, and then we computed the mean of individual correlation coefficients.

Finally, we determined the number of subjects that gave the same value to all the health states, identifying separately the number of subjects were not willing to accept any lifetime to improve quality of life (invariant, utility = 1). This the extreme form of insensitivity to health status severity in utility assessments.

2.8. Analysis of the robustness of the models

Predictive ability of the estimated models was tested by computing the Mean Absolute Error (MAE) and the Pearson correlation coefficient between the actual values of the 24 states directly provided by the respondents and the utilities predicted by the models. In addition, we analyzed, at the aggregate level, if the relative importance of the dimensions derived from the models is similar to that one directly provided by each participant. We obtained the percentage of participants who located each of the dimensions in the first or second place in order of seriousness, as a proxy of its importance. We rank ordered the dimensions according to this percentage and compared this ranking with that obtained from the estimated models. The last ranking is inferred from the relative weight of each attribute obtained by dividing its range, i.e. the difference between the highest and lowest coefficient, between the sum of the ranges of all attributes.

3. Results

3.1. Sample

Table 3 shows the main characteristics of respondents and the general population from Galicia. The sample was roughly representative of the Galician general population in terms of age and sex. Compared with the Galician population, our sample has a larger proportion of people in higher educational levels (secondary and university studies) and there are less people with higher earnings.

Table 3: Characteristics of respondents by type of questionnaire (%)

		Sample (n=312)	Population
Sex (female) [1]		47.4	51.6
Age[1]		41.5	45.4
Education[2]	Primary studies or less	37.5	54.0
	Secondary	39.4	27.9
	University	23.1	18.1
Habitat[1]	Rural	31.4	31.0
	Intermediate	31.1	33.3
	Urban	37.5	35.8
Live Alone[2]		13.5	19.6
Labour status[3]	Employed	59.8	46.9
	Pensioner/retired	10.9	23.8
	Unemployed	16.0	8.2
	Student	5.1	6.1
	Domestic tasks	8.3	9.9
Home income[2] (€ monthly)	<=500	5.9	3.7
	500-1000	13.2	19.1
	1000-1500	30.5	18.5
	1500-2000	25.7	16.4
	2000-3000	16.9	24.5
	>3000	7.7	17.9
Good health	(EQ-5D=11111)	76.3	
Close dependent	Any close dependent	53.2	
	Close dep. (not live together)	40.1	
	Close dep. (live together)	6.7	
Duration of interview (minutes)		22.5	
% Participants who placed this attribute in first or second place	Eat	24,99	
	Incontinence	37,23	
	Personal care	26,25	
	Mobility	28,89	
	Housework	3,18	
	Cognitive/mental	79,46	

The population data were obtained from:
[1] Census record (2011)
[2] Living conditions of Galician families survey (2007)
[3] Active population survey (2010). Office for National Statistics

3.2. Direct dependency state valuations

Table 4 shows mean utilities (and standard deviations) for the 24 states directly valued. For the overall sample (left columns), each dependency state was valued on average by 78 individuals, ranging from a minimum of 75 subjects to a maximum of 82. Most mean values were negative (14 out of 24), revealing that on average were regarded as worse than death. This is not a surprising result bearing in mind that the states used in this study describe, in general, more severe problems (dependence situations) than those usually described by means of generic

HRQoL instruments. Standard deviations are similar to those reported for severe EQ-5D and SF-6D health states (Badía et al., 2001; Dolan, 1997; Brazier et al., 2002; Lamers et al.., 2006).

Table 4 : Mean utilities for 24 states directly evaluated

	All participants			Consistent participants		
State	n	Mean utility	Std. Dev.	n	Mean utility	Std. Dev.
122222	82	0.17	0.54	77	0.18	0.53
133334	82	-0.45	0.47	77	-0.49	0.45
211121	82	0.58	0.47	77	0.60	0.46
214232	82	0.06	0.59	77	0.04	0.60
313331	82	0.17	0.60	77	0.17	0.61
323433	82	-0.49	0.51	77	-0.53	0.48
111221	78	0.65	0.42	77	0.66	0.42
112132	78	0.35	0.57	77	0.35	0.57
112211	78	0.60	0.45	77	0.60	0.45
223234	78	-0.47	0.54	77	-0.48	0.53
234333	78	-0.42	0.51	77	-0.44	0.50
333122	78	-0.23	0.62	77	-0.24	0.61
111112	75	0.40	0.57	66	0.50	0.49
113233	75	-0.12	0.66	66	-0.12	0.66
213322	75	-0.07	0.64	66	-0.07	0.64
222131	75	0.24	0.60	66	0.26	0.60
234431	75	-0.37	0.62	66	-0.37	0.62
334234	75	-0.54	0.51	66	-0.55	0.51
123121	77	0.30	0.57	73	0.32	0.58
212223	77	-0.16	0.67	73	-0.15	0.68
233432	77	-0.45	0.58	73	-0.48	0.57
314434	77	-0.60	0.51	73	-0.62	0.50
324332	77	-0.32	0.61	73	-0.32	0.61
333231	77	-0.19	0.64	73	-0.21	0.65

Regarding consistency of the responses, it is easy to check that, at the aggregate level, dominant profiles (indicating a lower dependency degree) always have a higher score. This is the case, for example, of state 211121 in comparison to states 214232, 313331 and 323433 (0.58 vs 0.06, 0.17, and -0.49, respectively). The picture is different at the individual level, however. There were 247 participants (79% of the respondents) that fulfilled all the dominance tests whereas 46 individuals (15%) failed only one test.

Correlation between the direct ranking of states and the ranking derived from TTO values (convergent validity) was larger for those participants that did not fail any dominance test (Spearman´s rho=0.73) or, at most, only failed one (Spearman´s rho=0.69), in comparison to those who failed twice or more times (Spearman´s rho=0.24). There were 20 participants who gave the same utility to all health states. However no participant in our survey refused to tradeoff life years in order to improve HRQoL, hence there was no subject in our sample who gave a utility of 1 for all health states.

As previously recognized (Devlin et al., 2003; Bravata et al., 2005) there is no consensus among researchers regarding how to handle utilities from inconsistent respondents. In our opinion, we think that it is necessary to distinguish between systematic departures from dominance and random error. In this vein, just subjects who failed only one dominance test at most would be included in the estimation of the tariff (Ryan et al., 2006). This is congruent, for example, with a 'trembling hand' notion of error (Harless and Camerer, 2004). Notwithstanding, a potential problem with dropping the remaining subjects from the analysis could be that their exclusion might alter the mean utility for a health state, affecting to the representativeness of the estimated tariff (Lamers et al., 2006). However, as Table 4 shows (right columns) this does not occur with our data. After excluding those respondents who made two or more inconsistencies (19 subjects; 6% out of the sample), each state was valued between 66 and 77 times, being mean utilities extremely similar to those obtained for the entire sample. Thus we did not include respondents who failed more than one dominance test in the regressions to estimate the DEP-6D algorithm.

3.3. DEP-6D models

Table 5 shows the estimates of two RE models estimated only with those participants that failed one dominance test at most. Model 1 includes all the regressors, even those (levels 3 and 4 of 'personal care' dimension) that lead to an inconsistent result because the disutility of the later (to be incapable of carrying out personal care) is smaller than that of the former (to need help for most personal care activities). To avoid this inconsistency, model 2 is constructed by aggregating PER3 and PER4, whose coefficients are not significantly different from each other according to the Wald statistic. Therefore, model 2 is a more 'efficient' model than model 1, reason why it is our preferred model. Although we tested both models with interaction terms similar to LEAST and MOST terms in the SF-6D, none of them was significant, so our algorithms reflect main effects only. In addition we re-estimated model 2 (not shown) including a series of dummy variables to account for the characteristics of the respondents. The coefficients of the variables describing the health states were not affected. From the added variables, only to live in a village with more than 10.000 inhabitants had a positive and significant coefficient.

Table 5: DEP-6D models

	model 1		model 2	
	Coefficient	p-value	Coefficient	p-value
Constant	0.776	0.000	0.773	0.000
EAT2	-0.152	0.000	-0.146	0.000
EAT3	-0.195	0.000	-0.188	0.000
INC2	-0.130	0.000	-0.131	0.000
INC3	-0.263	0.000	-0.263	0.000
PER2	-0.129	0.001	-0.133	0.001
PER3	-0.256	0.000		
PER4	-0.230	0.000		
PER3+4			-0.258	0.000
MOB2	-0.090	0.002	-0.086	0.003
MOB3	-0.133	0.000	-0.126	0.000
MOB4	-0.294	0.000	-0.289	0.000
HOU2	-0.066	0.140	-0.069	0.126
HOU3	-0.093	0.055	-0.089	0.066
MEN2	-0.228	0.000	-0.224	0.000
MEN3	-0.403	0.000	-0.403	0.000
MEN4	-0.527	0.000	-0.523	0.000
N	*293*		*293*	
Observations	*1758*		*1758*	

All the coefficients have the expected sign and are consistent estimates except for level 4 of the ‘personal care’ dimension in the model 1. The coefficients for model 2 show that the greatest decrements to health state value associated to the worst level in a dimension concern ‘mental problems’, ‘mobility’, ‘incontinence’, ‘personal care’, ‘eat’, and ‘housework’, in that order. These results are broadly consistent with the relative importance of the dimensions directly provided by the participants. In this way (see bottom side of Table 3) 79% of respondents place mental impairment in first or second place in order of severity, followed by incontinence (37%), mobility (29%), personal care (26%), eat (25%) and housework (3%). Only mobility and incontinence interchange their positions with regard to the ranking derived from parameter estimates.

The level of agreement between actual and predicted valuations for the 24 dependency states used to estimate model 2 is quite high both according to Pearson correlation coefficient (rho = 0.99) and the MAE (0.048). These results are quite similar to those reported for the EQ-5D (Dolan, 1997; Tsuchiya et al., 2002; Lamers et al., 2006) and even some better than those reported for the SF-6D (Brazier et al., 2002; Brazier and Roberts, 2004; Abellan-Perpiñan et al, 2012).

4. Discussion

This paper reports the estimation of a preference-based scoring algorithm for a new dependency health state classification system coined as DEP-6D. There are many dependency scales (e.g. Katz index of independence in Activities of Daily Living) but they cannot be used in CUA because they are not based on preferences. In most conventional scales assessment of functional status relies on simple counts of ADL dependencies. On the contrary, the model estimated for the DEP-6D generates preference weights for different levels of severity, allowing the estimation of a wide range of health states of dependency.

Other researchers before us (Bravata et al., 2005; Sims et al., 2005; Sims et al., 2008) have elicited preference weights for combinations of ADL dependencies. There are some remarkable differences between those studies and ours, however. An obvious difference deals with the way dependencies are characterized. Whereas our classification system resembles the combination of dimensions and different severity levels usual in generic HRQoL measures, such as the EQ-5D, SF-6D or HUI, each of the attributes used by the abovementioned authors (e.g. walking) contains one single level, reflecting just dependence or absence of it, but not how serious the dependence is in each dimension. In some respect, the 6 Katz ADLs plus an additional ADL of walking first used by Goldstein et al. (2002), and afterwards by Bravata et al. (2005) and by Sims and colleagues, are similar to the lowest levels of the first four dimensions of the DEP-6D. Hence the DEP-6D is able to describe a wider and richer set of dependence situations.

Besides, the DEP-6D includes cognitive impairment as one of its attributes, one critical dimension strongly associated with dependency (Andersen et al., 2004). Almost 80% of respondents in our sample regarded 'mental/cognitive problems' as the most severe dimension. Likewise coefficients estimated for mental impairments reflected the largest losses of utility.

Another difference concerns the source of preferences. The DEP-6D is fully consistent with the so-called 'societal perspective' (Drummond et al., 2005), according to which economic evaluations should include all potential effects and costs regardless of payer or beneficiary. As Gold et al. (1996) claim, a logical extension of that reasoning suggests that society's preferences should be gathered from a representative sample of general population. The DEP-6D algorithm is based on community preferences, not on preferences elicited from a specific sample of adults (aged 65 years or older).

Lastly, other differences arise from the analysis of the consistency and invariance of responses. Only 33% of participants in Bravata et al.'s (2005) study were not invariant and had no order inconsistencies. Moreover 19% of the respondents gave a utility of 1 for all health states of

ADL dependence. In our study no participant was invariant and just 6% of the sample had two or more inconsistencies. Extreme invariance explains the absence of variability in the mean utilities reported by Bravata et al. which range from 0.76 to 0.89. To explain these findings, the authors argue that it is possible that at least some of the invariant subjects did not understand actually the complexity of the valuation task. Factors such as low educational level or the high age of respondents (average age was 73.2) may play a role here.

Notwithstanding, in our opinion, apart from individual characteristics, there is another additional explanation for the high percentage of invariant responses reported by Bravata et al. (2005). The elicitation procedure used in their study was the standard gamble (SG) method. Nowadays it is well established (Bleichrodt, 2002) that the SG suffers from upward biases (particularly loss aversion) that generally will lead to overestimate the utility of a health state. There is a big amount of empirical evidence supporting the presence of upward biases in SG assessments (Bleichrodt et al., 2001, 2007; Pinto and Abellán, 2005; Abellán-Perpiñán et al., 2012). This tendency to overestimate the utility of health states could be exacerbated in Bravata et al.'s (2005) study because SG measurements were performed on a scale from cure to death instead of onto a full health-death scale. If the individual takes the health state of dependency to be valued, say Q, as her reference point, she trades off the gain from Q to be cured with probability p against the loss from Q to death with probability 1-p when she answers the SG. Hence, assuming that losses loom larger than gains, the probability of cure p required to offset the loss from Q to death with probability 1-p will be higher than the probability q that would be required if the gain was from Q to full heath. This follows from the fact that for $p=q$ the size of the gain from Q to be cured is lower than the size of the gain from Q to full health, since reaching full health is a better condition that simply be cured. Therefore the use of the SG, firstly, and the particular form it was administered, secondly, can also explain that respondents are not sensitive enough to the severity of the health status to be valued. This can contribute to explain that some subjects gave utility of 1 to all health states.

The elicitation technique used in our study was the TTO. There is some evidence (Bleichrodt and Johanesson, 1997; Abellán et al., 2009) that the TTO is more consistent with individual preferences than SG utilities. This higher external validity can be explained, from a theoretical point of view, because the biases for TTO measurements go in mutually opposite directions and tend to offset each other. Hence, overall, the TTO utilities are less biased than SG ones. For this reason Wakker (2008) claims that "Of the presently popular methods for measuring quality of life, the TTO is probably the best". Thus the use of the TTO instead of the SG can have helped in our study to lower error/objection responses.

As obvious the study introduced in this article has also limitations. One limitation comes from the fact that our design did not allow us to test for interaction effects. This means that our estimations assume that the six attributes of the DEP-6D are additively independent. Whereas we are aware that this is a restrictive assumption, it is also true that, indeed, the majority of the algorithms estimated for the EQ-5D and the SF-6D only reflects main effects (Tsuchiya et al., 2002; Lam et al., 2008; Ferreira et al., 2008; Brazier et al., 2009; Abellan-Perpiñan et al., 2012) or, at best, include some extreme level interaction term such as the intercept dummy N3 in the EQ-5D or the MOST/LEAST terms in the SF-6D (Dolan, 1997; Brazier et al., 2002; Brazier and Roberts, 2004). The predictive ability and consistency of our estimates is not worse than those reported for the mentioned multiattribute utility measures.

Otherwise, that the TTO utilities can be less biased than those measured with other methods does not mean that our TTO measurements cannot be improvable. We used an up-down procedure to reach the indifference between the two alternatives confronted in our TTO assessments. Once the indifference point was bounded between two durations (e.g. 7 and 8 years) we recorded it as the midpoint between both (e.g. 7.5 years). This means that utilities were calculated with an accuracy of 0.05 points of utility. A more accurate way to find the indifference point perhaps had affected our estimations, making, for example, that level 2 of 'housework' dimension ("Needs daily help") was significant different from the level 1 ("Does not need help").

To conclude, let us to note that although the DEP-6D shares some features common to the EQ-5D, namely, the method used to assess directly the health states (the TTO), and the econometric model applied to estimate the value set is (random effects), it should be clear that the health state classification system is quite different. In general DEP-6D health states are more severe conditions that EQ-5D health states. This is reflected in the utility range of both tariffs. In this way, the minimum value for the worst DEP-6D condition is -0.837 whereas the lowest utility for the TTO Spanish EQ-5D (Badía et al., 2001) is -0.654. This gap is larger if the comparison is done with respect to the Spanish SF-6D (Abellan-Perpiñan et al., 2012) with a minimum utility of -0.357. These discrepancies regarding the weights predicted by generic preference based measures enhance the relevance of having an instrument sensitive enough to describe dependence situations, able to yield, nevertheless, 'generic' QALYs. This is the case of the new tool introduced in this article, the DEP-6D.

References

Abellan-Perpiñan JM, Bleichrodt H, Pinto-Prades JL. The predictive validity of prospect theory versus expected utility in health utility measurement. Journal of Health Economics 2009; 28(6); 1039-1047.

Abellán-Perpiñán JM, Sánchez FI, Martínez JE, Méndez I. Lowering the 'floor' of the SF-6D Scoring algorithm using a lottery equivalent method. Health Economics 2012; 21; 1271-1285.

Andersen CK, Wittrup-Jensen KU, Lolk A, Andersen K, Kragh-Sorensen P. Ability to perform activities of daily living is the main factor affecting quality of life un patients with dementia. Health and Quality of Life Outcomes 2004; 2: 52.

Badia X, Roset M, Herdman M. Inconsistent responses in three preference-elicitation methods for health states. Social Science and Medicine 1999; 49; 943-50.

Badia X, Roset M, Herdman M, Kind P. A Comparison of United Kingdom and Spanish General Population Time Trade-off Values for EQ-5D Health States. Medical Decision Making 2001; 21; 7-16.

Beattie J, Covey J, Dolan P, Hopkins L, Jones-Lee M, Loomes G, Pidgeon N, Robinson A, Spencer A. On the contingent valuation of safety and the safety of contingent valuation: part 1-caveat investigator. Journal of Risk and Uncertainty 1998; 17(1); 5–26.

Bleichrodt H, Johannesson M. The Validity of QALYs: An Empirical Test of Constant Proportional Tradeoff and Utility Independence. Medical Decision Making 1997; 17; 21-32.

Bleichrodt H, Pinto JL, Wakker PP. Using Descriptive Findings of Prospect Theory to Improve the Prescriptive Use of Expected Utility. Management Science 2001; 47; 1498-1514.

Bleichrodt H. A New Explanation for the Difference Between Standard Gamble and Time Trade-Off Utilities. Health Economics 2002; 11; 447-456.

Bleichrodt H, Abellan-Perpiñan JM, Pinto-Prades JL, Mendez-Martinez I. Resolving inconsistencies in utility measurement under risk: tests of generalizations of expected utility. Management Science 2007; 53, 469–482.

Bravata DM, Nelson LM, Garber AM, Goldstein MK. Invariance and inconsistency in utility ratings. Medical Decision Making 2005; 25(2);158-67

Brazier J, Deverill M. A checklist for judging preference-based measures of health related quality of life: Learning from psychometrics. Health Economics 1999; 8(1); 41-51.

Brazier J, Roberts J, Deverill M. The estimation of a preference-based measure of health from the SF-36. Journal of Health Economics 2002; 21; 271-92.

Brazier J, Roberts J. The estimation of a preference-based measure of health from the SF-12. Medical Care 2004; 42; 851-9.

Brazier J, Fukuhara S, Roberts J, Kharroubi S, Yamamoto Y, Ikeda S, Doherty J, Kurokawa K. Estimating a preference-based index form the Japanese SF-36. Journal of Clinical Epidemiology 2009; 62; 1323–1331.

Burge P, Netten A, Gallo F. Estimating the value of social care. Journal Health Economics 2010; 29(6); 883-94

Chilsholm, D., Healey, A., Knapp, M. QALYs and mental health care. Social Psychiatry and Psychiatric Epidemiology 1997; 32; 68-75.

Coast J, Flynn T, Natarajan L, Sproston K, Lewis J, Louviere J, Peters T. Social Valuing the ICECAP capability index for older people. Social Science and Medicine 2008; 67 (5); 874–882.

Devlin NJ, Hansen P, Kind P, WilliamsA. Logical inconsistencies in survey respondents' health state valuations: a methodological challenge for estimating social tariffs. Health Economics 2003; 12; 529–44.

Dolan P. Modelling valuations for EuroQol health states. Medical Care 1997; 35; 1095-108.

Dolan P, Kind P. Inconsistency and health state valuations. Social Science and Medicine 1996; 42; 609–15.

Donaldson C, Atkinson A, Bond J, Wright K. Should QALYs be programme-specific? Journal of Health Economics 1988; 7; 239-257.

Drummond M, Sculpher M, Torrance G, O'Brien B, Stoddart G. Methods for the economic evaluation of health care programmes. Oxford: Oxford University Press; 2005.

Fedorov VV. Theory of optimal experiments. Academic press: New York; 1972.

Feeny D, Furlong W, Torrance GW, Goldsmith CH, Zenglong Z, Depauw S, Denton M, Boyle M. Multi-attribute and single-attribute utility function for the Health Utility Index Mark 3 system. Medical Care 2002; 40; 113-28.

Ferreira PL, Ferreira LN, Pereira LN. How consistent are health utility values? Quality of Life Research 2008; 17; 1031–1042.

Gold MR, Siegel JE, Russell LB, Weinstein MC. Cost-effectiveness in health and medicine. New York: Oxford University Press; 1996.

Goldstein MK, MillerDE, Davies S, GarberAM. Quality of life assessment software for computer-inexperienced older dults: multimedia utility elicitation for activities of daily living. Proc AMIA Symp. 2002; 295–9.

Harless D, Camerer CF. The Predictive Utility of Generalized Expected Utility Theories. Econometrica 1994; 62; 1251-1289.

Kane R, Kane R. Assessing Older Persons Measures, Meaning, and Practical Applications. New York: Oxford University Press; 2002.

Katz S, Ford A, Moskowitz R, Jackson B, Jaffe M. Studies of Illness in the Aged. The Index of ADL: A Standardized Measure of Biological and Psychosocial Function. JAMA 1963; 185(12); 914-919

Keeney RL, Raiffa H. Decisions with multiple objectives–preferences and value tradeoffs, Cambridge & New York: Cambridge University Press; 1993.

Lam CL, Brazier J, McGhee SM. Valuation of the SF-6D health states is feasible, acceptable, reliable, and valid in a Chinese population. Value in Health 2008; 11; 295–303.

Lamers LM, Stalmeier PF, Krabbe PF, Busschbach JJV. Inconsistencies in TTO and VAS values for EQ-5D health states. Medical Decision Making 2006; 26 (2); 173-81.

Olsen JA, Donaldson C, Shackley P, and the EuroWill Group. Implicit versus explicit ranking: on inferring ordinal preferences for health care programmes based on differences in willingness-to-pay. Journal of Health Economics 2005; 24; 990–996.

Patrick DK, Starks HE, Cain KC, Uhlmann RF, Pearlman RA. Measuring preferences for health states worse than death. Medical Decision Making 1994; 14; 9-18.

Pinto JL, Abellán-Perpiñán JM. Measuring the health of populations: the veil of ignorance approach. Health Economics 2005; 14; 69 – 82.

Ryan M, Netten A, Skatun D, Smith P. Using discrete choice experiments to estimate a preference-based measure of outcome—An application to social care for older people. Journal of Health Economics 2006; 25; 927–944

Sims TL, Garber AM, Miller DE, Mahlow PT, Bravata DM, Goldstein MK. Multimedia Quality of Life Assessment: Advances with FLAIR. Proc AMIA Symp 2005; 694–698.

Sims T, Holmes TH, Bravata DM, Garber AM, Nelson LM, Goldstein MK. Simple counts of ADL dependencies do not adequately reflect older adults' preferences toward states of functional impairment. Journal of Clinical Epidemiology 2008; 61(12); 1261–1270.

Torrance GW, Thomas W, Sackett D. A utility maximisation model for evaluation of health care programmes. Health Service Research 1972; 7(2); 118-33.

Torrance GW. Multi-attribute utility theory as a method of measuring social preferences for health states in long-term care. In Kane RL, Kane RA (eds.): Values in long-term care. Lexington Books, DC Health, Lexington, MA; 1982.

Torrance G W. Measurement of health-state utilities for economic appraisal: A review. Journal of Health Economics; 1986; 5; 1-30.

Torrance GW, Feeny D, Furlong W, Barr R, Zhang Y, Wang Q. Multiattribute preference functions for a comprehensive health status classification system: health utilities index mark II. Medical Care 1996; 34; 702–2.

Tsuchiya A, Ikeda S, Ikegami N, Fukuda T, Hamashima C, Hisashige A, Tamura M. Estimating an EQ-5D population value set: the case of Japan. Health Economics. 2002; 11; 341–53.

Wakker PP. Lessons learned by (from?) an economist working in medical decision making. Medical Decision Making 2008; 28; 690-698.

Witternberg E, Prosser L. Ordering errors, objections and invariance in utility survey responses. Applied Health Economics and Health Policy 2011; 9(4): 225-241.