Valuing Health States using the MACBETH non-numerical approach

Andreia Cristina Sanganha Agostinho

Mestrado Integrado em Engenharia Biomédica

Instituto Superior Técnico, Universidade de Lisboa, Portugal

Thesis supervised by Professors Mónica Oliveira (DEG-IST) and Paulo Nicola (IMPSP-FMUL)

Abstract: *Quality-adjusted life years (QALYs), a measure commonly used in health technology appraisals, incorporates in a summary index changes in quantity and in quality of life. The use of QALYs requires the modelling of individuals' quality of life for distinct health states into utility-based QALY scores, with preference elicitation methods such as the Visual Analogue Scale, the Time Trade-Off (TTO) and the Standard Gamble. Given that these numerical elicitation methods may be associated with a high cognitive effort and their results may be differentiated by individual's and population's numeracy, in this study we explore the use of the MACBETH (Measuring Attractiveness by a Categorical Based Evaluation Technique) non-numerical approach to evaluate health states.*

A new protocol for preference elicitation based on the MACBETH approach (only requiring qualitative judgments) was developed and tested in a web survey alongside with TTO in a sample of the Portuguese general population (n=243). The web-survey asked individuals to value directly 25 different EQ-5D health states with the two methods, as well as for their preferred method. Individuals' numeracy was also assessed.

As key results, we found out that: the mean values derived from MACBETH and TTO are strongly correlated (Pearson r = 0.962); participants with higher and lower levels of numeracy preferred expressing value judgments with MACBETH; and mean values obtained differed on the basis of numeracy level.

Results suggest that it is worth considering the use of non-numerical preference elicitation methods in health. Further research may be directed to understand eventual advantages of non-numerical methods in special disease populations, diseases states and across population from different countries.

Keywords: QALYs; preference-based instruments; health states valuation; MACBETH; TTO

1. Introduction

In a context where resources for the provision of health care are limited and decisions have to be made about how they are allocated, the field of health technology assessment (HTA) has shown a remarkable growth. [1] HTA is a multidisciplinary science that aims at bringing together evidence with decision-making to help policymakers, clinicians and patients to understand the relative value of technologies and has been evolving towards the maximization of effectiveness with available resources. [2]

Within HTA, different techniques of economic evaluation have been widely used to compare two or more alternative courses of action in terms of both costs and health benefits: cost-benefit analysis (CBA), cost-effectiveness analysis (CEA) and cost-utility analysis (CUA). These different types of economic evaluation are defined by the way in which outcomes are measured, thus CBA assigns monetary value to health outcomes, CEA uses clinical outcomes in natural unidimensional units and CUA, as a type CEA, focuses on a single summary measure as quality adjusted life years (QALYs) which is based on the patients' preferences for being in a particular health-state. The use of QALYs has been of particular interest since it enables comparisons across a wide range of alternatives which produce multiple results within different areas of healthcare and allows the incorporation of patients' preferences into the decision-making process. In fact the QALY is the health outcome measure recommended by the NICE (National Institute for Health and Clinical Excellence) in UK for technology appraisals. [3]

The QALY is a single measure of health outcome which simultaneously incorporate the impact on both length and quality of life; it is defined as the product of the patient's life expectancy and the quality of life in those remaining years. Changes in quantity of life, expressed in terms of survival or life expectancy, are measured in years and the quality of life adjustments are based on a set of preference values or weights called utilities, one for each health state, in which preference can be equated with value or desirability. [4][5] Utilities in QALYs are measured on an interval scale, where 1 refers to full health and 0 refers to death. Some severe health states may be considered as

being worse than death, resulting in negative utilities. [6][7]

Three main conventional methods have been used for eliciting preferences for health states: Visual Analogue Scale (VAS), Time Trade-Off (TTO) and Standard Gamble (SG). Different methods have been shown to lead to different results and there has been no consensus towards the most preferred method, since each of them has distinct pros and cons. [8][9]

Given that these numerical elicitation methods are associated with a high cognitive effort and complex administration modes, and recent studies determined that their results may be influenced by individual's and population's numeracy, in this study we explored the use of the MACBETH (Measuring Attractiveness by a Categorical Based Evaluation Technique) non-numerical approach to evaluate health states.

2. Background and Literature Review

Health-related quality of life (HRQoL) is the most established patient reported outcome (PRO) measure and several types of instruments have been designed for measuring HRQoL which can be classified into three groups: disease specific instruments, generic health profiles and preference-based measures. [10]

The use of QALYs has been increasing throughout the world and that leads to a growing interest in preference-based measures of HRQoL that provide a single summary score (HRQoL weight) derived from preferences of the general population. [10] There are two main approaches to measuring preference-based scores: direct measurements by means of techniques such VAS, SG and TTO; and indirect determination instruments also known as multiattribute health status classification systems with preference scores. [6]

The VAS is the simplest method but this is not a choice-based technique and has the least grounding in economic theory. [4][11] However its simplicity leads to important advantages in empirical performance and it is a useful tool used as a "warm-up" exercise before other methods.

The SG is the classical method of measuring cardinal preferences, since it is based directly on the axioms of utility theory. [6] The SG approach involves a choice with an element of risk and uncertainty in decisions faced by respondents. [12]

The TTO approach like SG is a choice-based technique. The choice involves asking respondents to consider the number of life years they would be willing to sacrifice to avoid a certain poorer health state. [11][6]

The two most widely used methods to measure patients' health preferences (values or utilities, respectively) are TTO and SG. However there are theoretical and empirical drawbacks associated with these methods: responses are likely to be influenced by factors such as risk behavior of the respondents, time preference or aversion to loss; protocols are complex and demand a high cognitive effort; different tasks for states better and worse than death; and potential violation of underlying assumptions. [7][13][14]

To overcome complex and time-consuming direct measurement of preferences, unrealistic in clinical research, some indirect preference-based measures of health status have been developed to facilitate the determination of preference scores. These instruments consist in a descriptive system in which the health status of the individual is classified and that provides a preference-based score based on a scoring formula that comes with the system and is based on directly measured preferences of the general public. [15] Examples are the Quality of Well-Being (QWB), Health Utilities Index (HUI), EuroQol-5D (EQ-5D) and Short-Form 6D (SF-6D). [4] These instruments differ in terms of the health dimensions that are included, the number and description of levels defined for each dimension, the population on which the preferences are based and in terms of the valuation method: TTO was used to value the EQ-5D and SG is used to value HUI and SF-6D. [4]

In addition to the direct measurements techniques, discussed above, there is an ongoing research in the use of different methods for the elicitation of health state values, in particular ordinal methods such as discrete choice experiments (DCEs) and ranking exercises that offer advantages relative to ease of comprehension and administration and reduced cognitive burden which are particularly important in settings with limited educational attainment and low numeracy level. [14] The latter point has been discussed in recent studies that have shown that patients' quantitative skills (i.e., numeracy) may influence values obtained from conventional elicitation preferences techniques that are inherently quantitative and also individuals' preferred mode of expressing value judgments, in numbers or words. [16][17][18]

The present work is motivated by the perceived limitations of the conventional valuation techniques, in particular a high cognitive effort and a complex protocol administration, associated with recent results that indicated that preference values derived from numerical techniques may be differentiated by individual's and population's numeracy. In this context we explore the use of the MACBETH non-

numerical approach to evaluate health states (as will be described in section 3.1). A new protocol is designed and tested in a web survey alongside with a conventional protocol, in this case TTO, to allow for comparison between both methods. Besides that, this study also wants to investigate some questions related to the impact of individuals' numeracy: 1) will numeracy affect individuals' preferences between numerical and non-numerical elicitation preferences techniques? [16]; 2) will population's numeracy influence obtained values? [17][18]; and 3) will numeracy increase consistency in health state evaluations (obtained from different techniques) ? [18]

3. Methodological Framework

According to the previously defined objectives, the proposed methodology involves a set of progressive steps that make use of different tools. The main steps of the proposed methodology are: design a new protocol based on the MACBETH non-numerical approach to valuing health states; develop and implement a web survey to test the new protocol alongside with TTO and simultaneously collect data about preferred mode of expressing value judgments and apply a numeracy test; valuing health states based on collected value judgments and identify exclusion criteria; and at last analyze obtained results.

In the next sections these steps of the proposed methodology will be detailed described.

3.1. MACBETH approach and new protocol to valuing health states

MACBETH (Measuring Attractiveness by a Category-Based Evaluation Technique) is a non-numerical approach widely used to support decision-making that allows measurement of the relative value of different options through a non-numerical pairwise comparison of differences in attractiveness between options based on seven qualitative categories: no difference, very weak, weak, moderate, strong, very strong or extreme. [19]

The adoption of the MACBETH approach for the purpose of valuing health states was primary motivated by the reduced cognitive burden of this type of task, that does not require quantitative judgments and for this reason is not influenced by numerical skills.

The MACBETH questioning, to valuing health states, requires qualitative judgments of differences in attractiveness obtained by a pairwise comparison between health states: given two health states with state x better than state y, the individuals are asked for differences in attractiveness between x and y. The elicited judgments are introduced in M-MACBETH software filling a matrix of judgments like Figure 1. [20]

	a=upper	b	c	d	e=lower
a=upper	no	very weak	moderate	positive	positive
b		no	weak-mod	moderate	positive
c			no	weak	positive
d				no	no
e=lower				no	no

extreme
v. strong
strong
moderate
weak
very weak
no

Figure 1- Consistent MACBETH matrix of judgments (options a, b, c, d and e correspond to different health states).

For a set with n health states, with states ordered by decreasing preferences, it is not necessary to perform all of the n(n-1)/2 paired comparisons and populate the upper triangular part of the matrix completely. The minimal number of judgments required is n-1, e.g., comparing one state with the remaining ones or comparing the states rank-ordered consecutively, however, it is recommended to ask for additional judgments to perform consistency checks. [19]

Each time a qualitative judgment is introduced in the matrix, M-MACBETH tests the consistency of all the judgments made and when inconsistency is detected suggestions are made to resolve it. [20] When a consistent matrix is obtained MACBETH uses a mathematical programming algorithm to derive scores for all the options from the set of quantitative judgments.

The MACBETH protocol for valuing heath states is defined and it consists basically of four different steps, according to the schematic representation of Figure 2: 1) Order the set of health states presented in terms of perceived attractiveness by individuals. In the context of valuing health states is necessary anchor the values obtained for different health states on a scale between perfect health and death, so in MACBETH protocol it is necessary to include the states "perfect health" and "immediate death" in the set of health states presented. These states are subsequently defined as upper reference and lower reference which are assigned values of 0 and 100, respectively. 2) Judge qualitatively the difference of attractiveness between different health states. In this task it is required judgments between health states rank-ordered consecutively and two additional judgments to check consistency. 3) Order and fill the M-MACBETH matrix of judgments according to data collected in previous steps. 4) Obtain a value scale with the different health states valued.

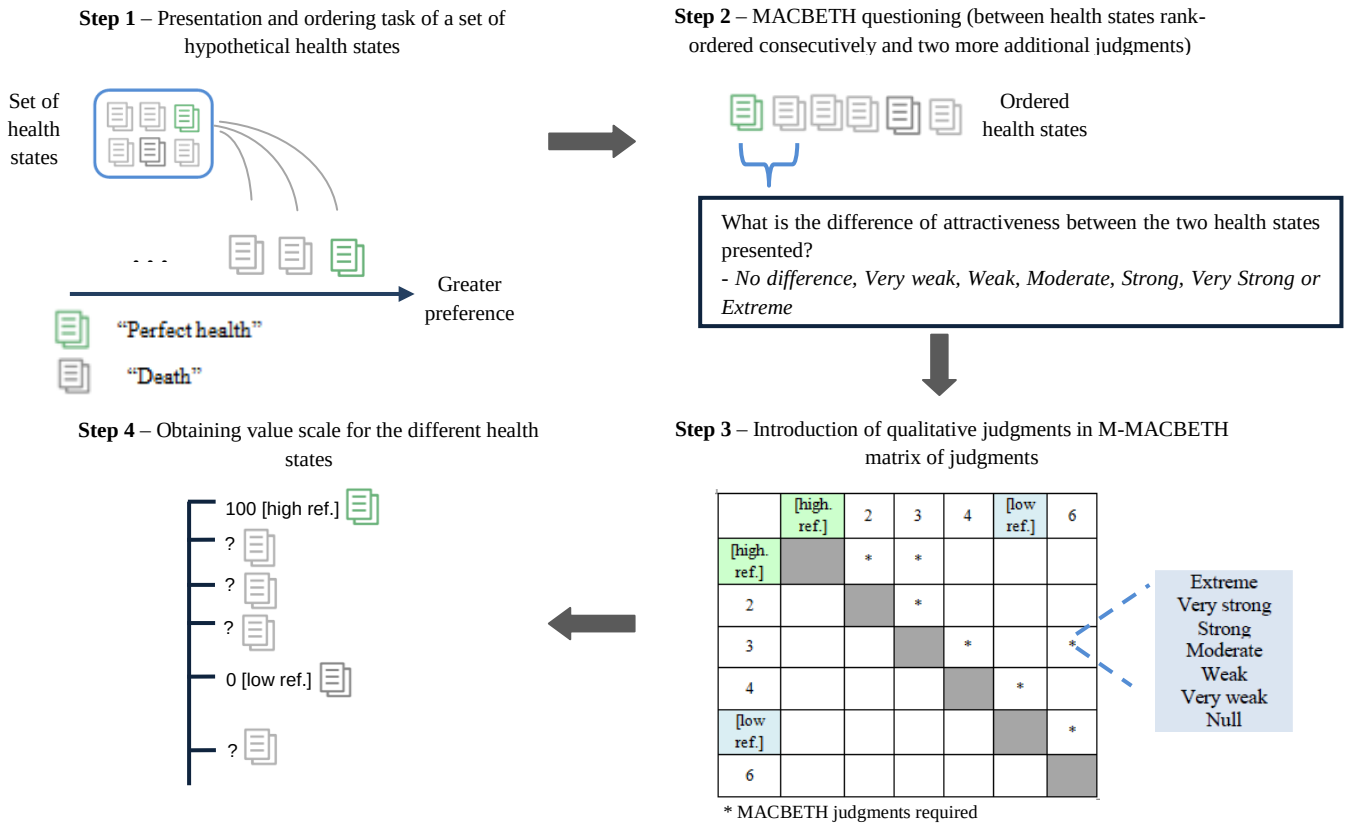

Figure 2- Schematic representation of the new protocol based on the MACBETH approach to valuing health states.

3.2. Web survey

The main components of the proposed methodology include the design of the survey, and then its online development and implementation using Qualtrics platform. The decision about the development of an online survey was motivated by its easy and quickly distribution and by the numerous logistical challenges and resource limitations associated with conventional methods for eliciting preference – face to face interviews. Given that increasingly more internet and new technologies are accessible to the majority of the population in developed countries this approach for research should be considered for future valuation studies. [21]

The purpose of this survey was to evaluate the new non-numerical protocol, compare the process with a conventional TTO and collect data about preferences between the two value elicitation techniques and individuals' numeracy. The main steps underlying the construction and implementation of this online survey are described below.

A. Selection of health states

Health states in the present study were defined in terms of EQ-5D descriptive system. This measure of HRQoL tends to be the most commonly used in CUA analysis and is the preferred instrument for NICE. [4] The choice for the EQ-5D descriptive system in this study merges with the choice of the conventional method and is highly motivated by a recent study that estimates the EQ-5D value set using TTO for Portugal. [22]

The EQ-5D descriptive system consists of five dimensions (mobility, self-care, usual activities, pain/discomfort and anxiety/depression) with three possible levels each (level 1 - no problems; level 2 - some problems; level 3 - extreme problems); thus defining 243 (3^5) heath states. Health states descriptions are constructed by taking one level of each attribute, e.g., 11111 represents the best and 33333 the worst state. [23]

In this study a set of 24 health states was chosen for valuation plus the health states 11111, 33333 and "immediate death", the same as in EQ-5D valuation in Portugal thus allowing result comparisons. [22] Since previous studies showed that respondents are not capable of valuing more than approximately 13 health states within the same exercise, the health states were divided into four equally sized groups according to their severity. [24] Each respondent was randomly assigned to one of those groups (Table 1).

Table 1- Health states set assignments.

Group 1	Group 2	Group 3	Group 4
13311 ***	12111 *	11113 **	21111 *
22222 ***	11131 **	32313 ***	23232 ****
11112 *	32211 ***	11211 *	11121 *
11133 **	21323 ***	22121 **	11312 **
32223 ****	22233 ****	13332 ****	33323 ****
33321 ****	23313 ****	33232 ****	22122 ***
33333	33333	33333	33333

(*very mild health state, **mild health state, ***moderate health state, ****severe health state)

B. Structure of the survey and elicitation tasks

Each respondent were asked to: 1) describe their own health using the EQ-5D instrument including the self-classifier and the EQ VAS; 2) directly value a set of hypothetical health states using the TTO and the MACBETH protocol (tasks presented in random order); 3) complete a numerical test composed by three validated questions from [17]; 4) identify their preferred way of expressing value judgments (numerical technique, non-numerical technique or indifferent); and 5) report their socio-demographic characteristics.

To value different hypothetical health states, the respondents were told that they had to picture themselves in each state for a period that would last 10 years, after which they would die.

TTO. The TTO method consists of two different tasks. First is requested to respondents to indicate whether the heath state being evaluated was better or worse than death. According to the first response one question was applied for states respondents valued as better than death and other for states valued as worse than death (see [22]).

The protocol followed in the present study resulted from an adjustment of the commonly used protocol in order to reduce the time consumed in this task. [24][25] The process has been simplified through the direct appearance of a horizontal scale, limited by values 0 and 10, representing the number of years in full health (state better than death) or the number of years in the target health state (state worse than death). The respondents are asked to directly indicate the indifference point between alternatives presented.

MACBETH. As referred above the MACBETH protocol also consists of two different tasks (Figure 2- Step1 and 2). First in a rank ordering exercise the respondents had all states (including death) in front of them and were asked to select the best state, in an iterative process were the selected state disappears, until an order is established. The second task consists in obtaining the qualitative judgments as described above.

C. Distribution strategy and target population

The target population for the study consisted of a sample of the Portuguese general population, aged 18 and over. Individuals were recruited to participate in this study through an invitation sent by email which contains a link for the survey and also a request to individuals forward the invitation to other contacts, thus following a non-probability sampling strategy –snowball sampling or networks sampling. The initial sample consisted in 54 individuals.

3.3. Valuation procedure

When data collection was concluded, the next step consisted in the derivation of values for the different health states, taking into account the numeric and non-numeric judgment gathered according to each method.

TTO. If perfect health and death are given the values 1 and 0, respectively, then the values for health states that were valued as better than death are given by the formula $h = t/10$ and for health states that were valued as worse than death the value is given by the formula $h = (-10 + t)/t$. In both formulas t represents the indifference point.

MACBETH. In the MACBETH approach the protocol defined does not differ for health states considered better or worse than death. First, full health (11111) and death are defined as upper and lower references, respectively and the health states under evaluation are introduced into the MACBETH matrix. Based on the rank ordering exercise the MACBETH matrix is ordered and then the qualitative judgments (word categories) elicited for differences of attractiveness between pairs of health states are introduced. If a consistent matrix of judgments is obtained a numerical scale of relative values is proposed. (Figure 2- Step3 and 4)

Given the defined approach to data collection, the gathering of qualitative judgments according to MACBETH protocol and their introduction into the M-MACBETH matrix occurred at different times, so it was not possible to present each individual with the scale proposed by M-MACBETH for validation; the proposed scale is, therefore, automatically validated.

Monotonic Transformation. For both MACBETH and TTO it is necessary to rescale the obtained

values in such a way that the scales vary between -1 and 1. In TTO since responses are measured between 0 and 10 with increments of 0.1, h ranges from -99 to 1. Thus, negative values were adjusted so that they were bounded by -1 and 0. In order to rescale these values the following monotonic transformation has been used: $h^{'} = h/(1-h)$; for consistency with previous valuation studies. [22] In MACBETH no theoretical lower boundary exists for states worse than death and since lower reference value defined is 0 and upper reference value is 100, in a first stage values are transformed by a linear transformation $h' = h/100$ and, then, a monotonic transformation is also applied for states worse than death (negative values).

3.4. Exclusion criteria

For analyzing the survey results, one expects that all respondents whose responses reflected inconsistencies caused by a lack of understand or misinterpretation of an exercise are excluded. Excluding respondents from the analysis can be separated into conditions which may affect numerical judgments and conditions which may affect non-numerical judgments. In the former category the following objective criteria were applied: 1) all states were valued worse than death; 2) all states given the same value; 3) for states better than death: $t = 0$; and 4) for states worse than death: $t = 0$. Additionally two subjective exclusion criteria based on "logical inconsistency" and "serious logical inconsistency", previously defined in other studies, were analyzed. [22] A logical inconsistency occurs, at respondent level, if one state of a pair is better than the other one at least one dimension and not worse in any other dimension, and the valuation for the former state is worse than the valuation for the latter case. It is a "serious logical inconsistency" if the difference in valuation was greater or equal to 0.5. [22]

In the category of exclusions associated with non-numerical judgments the following objective criteria were applied: 1) logical inconsistency resulting from the rank order exercise; 2) all states were valued worse than death; and 3) inconsistent MACBETH matrix.

When exclusions associated with numerical and non-numerical judgments were identified for the same individual, the total questionnaire was excluded; however when only one of the methods has one or more problems identified, only the responses associated with that method are excluded.

3.5. Data analysis

Data analysis consists in the final step of the proposed methodology and aims to: describe the process of obtaining a valid population sample for analysis; characterize the population sample in relation to sociodemographic variables, current health status and numeracy level; compare the numerical and the non-numerical elicitation techniques; find out individuals' preferred method; and finally study the influence of numeracy level (on preferences, on health states valuations and on the exclusion of answers).

Descriptive statics were calculated for different variables, including different sample's characteristics. Comparisons between subgroups were made using parametric tests (t tests and ANOVA) and non-parametric tests (χ^2 tests, fisher exact test and Mann-Whitney test). Correlation coefficients (Pearson and Spearman) were also included to compare values scales obtained. A multinomial logistic regression was applied to study the influence of numeracy in the preferences expressed. All statistical analysis were performed using R (v.3.1.3) and a 5% significance level was considered ($p<0.05$).

Additionally another type of analysis is performed that consists in the adjustment, at individual level, of the proposed MACBETH scale to the TTO scale obtained.

4. Results

4.1. Obtaining a valid sample

Figure 3 shows a flow diagram of the process of obtaining a valid sample population for analysis. A total of 348 individuals, invited by email, accepted to participate in this study but only 243 completed the survey. The completion rate is, approximately, 70%.

The total number of exclusions associated with both MACBETH and TTO, partial exclusions MACBETH and partial exclusions TTO was 33, 98 and 27, respectively. The main reasons for exclusions associated with MACBETH task were: inconsistent MACBETH matrix (43.6%) and logical inconsistency resulting from the rank order (8.6%). In relation to TTO task the main exclusions were for states better and worse than death with $t = 0$, respectively 7% and 20.2%. Additional criteria were analyzed: 51.9% of the TTO respondents shown logical inconsistency and 4.1% shown serious logical inconsistency.

A valid sample for analysis with 210 individuals was obtained: 112 (46%) valid judgments associated with MACBETH and 183 (75%) valid judgments associated with TTO; only 85 (40.5%) individuals have valid responses for both TTO and MACBETH.

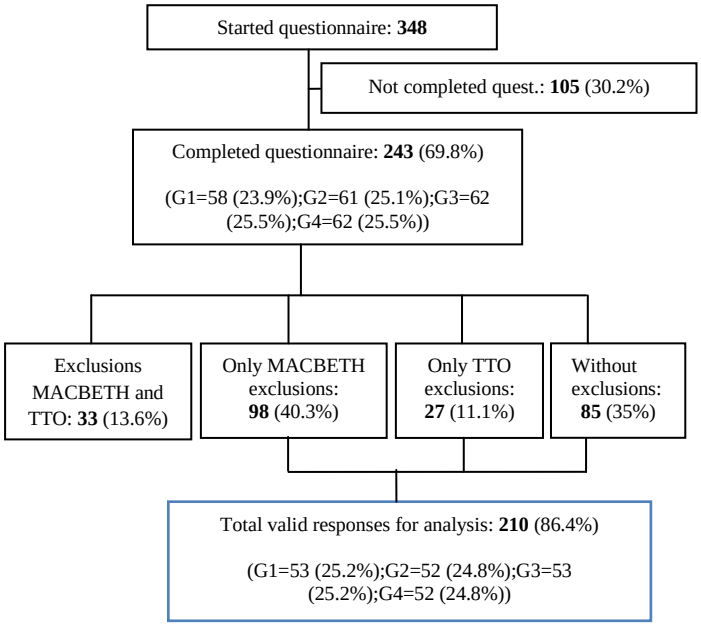

Figure 3- Flow diagram of the process of obtaining a valid sample population for analysis (G1, G2, G3, G4 – Set of responses obtained for each one of the four sets of health states).

Given that the order of presentation of the MACBETH and the TTO tasks was randomized, at individual level, it was considered of interest to analyze a possible influence in excluded and incomplete questionnaires. Regarding to the influence on the number of exclusions statistically significant differences were not found; however, it was found that for incomplete questionnaires a higher proportion of individuals completed the first task displayed if this was the MACBETH task.

4.2. Characteristics of the sample

Table 2 shows the main characteristics of the study sample. The sample includes a slight majority of women (63.8%). Age ranges from 18 to 74 with mean age of 34 years old (SD=13). The majority of the sample is single or married/living together as a couple. Almost 73.8% of the sample has a high educational level.

All respondents were asked to fill out the EQ-5D questionnaire expressing their own health states. In general all respondents placed themselves in very good health states, only in pain/discomfort and in anxiety/depression, 21.4% and 24.3% respectively reported moderate problems.

Table 2- Study sample characteristics.

		Valuation (n=243)	Valuation after exclusions (n=210)
Gender (%)	Female	64.0	63.8
	Male	36.0	36.2
Age (%)	Mean (SD)	34.4 (13.1)	34.3 (13.1)
	18-24	31.7	33.3
	25-30	18.9	17.6
	31-44	25.5	25.2
	> 44	23.9	23.8
Educational attainment (%)	Less than secondary	2.1	1.4
	Secondary	26.3	24.8
	High than secondary	71.6	73.8
Marital status (%)	Single	51.0	50.0
	Married/ living with a partner	45.7	48.1
	Divorced/separated	2.5	1.4
	Widowed	0.8	0.5
Occupational status (%)	Student	33.3	36.2
	Employed	53.9	51.4
	Unemployed	9.7	3.3
	Retired	4.1	4.3
	Domestic	1.2	1.4
	Other situation	3.7	3.3
Household (%)	1-2 elements	28.4	26.7
	3-4 elements	63.0	64.3
	5 or more elements	8.4	9.0
Chronic disease (%)	Yes	19.3	18.6
	No	77.8	79.1
	NA/DK	2.9	2.4
Numeracy (%)	Mean (SD)	2.4 (0.8)	2.4 (0.8)
	0 right answers	2.9	2.9
	1 right answer	11.5	11.4
	2 right answers	27.6	26.2
	3 right answers	58.0	59.5

Differences between population subsamples are investigated, in particular between the total sample and the sample after exclusions and between the MACBETH and the TTO included responses and excluded. No statistically significant differences at 5% level were found.

4.3. Health-states values

Descriptive statics for hypothetical health states, the number of valuations per health state, the percentage of negative valuations and the difference between mean values are reported in Table 3.

The mean health state value for MACBETH is 0.042 (SD=0.29) with a range between -0.080 (33333) and 0.860 (12111); for TTO is the mean value is 0.35 (SD=0.37) with a range between -0.446 (33333) and 0.831 (11121). For MACBETH the only health state valued as negative was 33333.

Table 3- Obtained values for 25 hypothetical health states.

	MACBETH			TTO			
State	**n**	**Mean±SE**	**% neg.**	**n**	**Mean±SE**	**% neg.**	**Dif.**
11112	26	0.85±0.03	0	48	0.81±0.03	0	0.04
12111	28	0.86±0.02	0	46	0.83±0.02	0	0.032
11211	31	0.84±0.02	0	42	0.80±0.03	0	0.038
21111	27	0.81±0.02	0	47	0.78±0.03	0	0.033
11121	27	0.85±0.02	0	47	0.83±0.03	0	0.023
11131	28	0.65±0.04	0	46	0.56±0.05	2.2	0.095
11113	31	0.58±0.07	10.8	42	0.56±0.06	7.1	0.012
11133	26	0.34±0.07	15.4	48	0.40±0.07	8.3	-0.062
22121	31	0.58±0.05	3.2	42	0.68±0.03	0	-0.095
11312	27	0.55±0.03	0	47	0.53±0.06	6.4	0.014
13311	26	0.43±0.05	7.7	48	0.50±0.04	2.1	-0.065
22122	27	0.51±0.04	0	47	0.59±0.04	0	-0.086
21323	28	0.35±0.04	3.6	46	0.36±0.05	2.2	-0.007
32211	28	0.40±0.05	3.6	46	0.26±0.05	8.7	0.134
32313	31	0.14±0.06	32.3	42	0.14±0.08	21.4	0.011
22222	26	0.58±0.04	0	48	0.46±0.05	4.2	0.119
33232	31	0.04±0.06	32.3	42	-0.21±0.07	50.0	0.253
23232	27	0.28±0.04	11.1	47	0.01±0.07	25.5	0.265
13332	31	0.17±0.06	29.0	42	0.01±0.07	26.2	0.161
22233	28	0.26±0.05	14.3	46	0.16±0.07	17.4	0.101
32223	26	0.11±0.06	34.6	48	0.06±0.07	16.7	0.045
33321	26	0.15±0.06	26.9	48	0.07±0.08	20.8	0.073
23313	28	0.29±0.04	3.6	46	0.18±0.06	13.0	0.114
33323	27	0.05±0.04	33.3	47	-0.27±0.08	51.1	0.315
33333	112	-0.08±0.02	48.2	183	-0.45±0.04	63.9	0.366
						MAD:	**0.1015**

SE – *Standard error*; **MAD** – *Mean Absolute Difference*;

Despite some differences, mean values do not differ remarkably between both methods. The absolute difference is greater than 0.1 for 9 health states (36%); and the absolute difference is greater than 0.05 for 15 health states (60%). Negative values are assigned to some severe health states and, in general, are consistent between MACBETH and TTO.

Figure 4 shows a comparison between obtained mean values for 25 hypothetical health states for TTO and for MACBETH, ordered by MACBETH values.

Pearson correlation coefficients were determined, r = 0.962 (p = 1.6e-14), which reveals a strong correlation between the two mean scales obtained. The determination coefficient, r^2 = 0.926, indicates that, approximately, 93% of the variability in one scale can be explained by the other.

From the test of the adjustment, at individual level, of the proposed MACBETH scale to the TTO scale, using a subsample that consists in Group 1, it was verified that the numerical values from TTO scale are not consistent with MACBETH qualitative judgments.

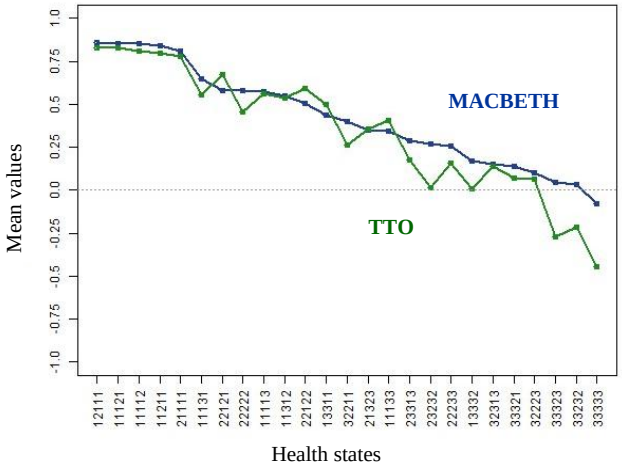

Figure 4- A comparison between obtained mean value for 25 hypothetical health states for TTO and for MACBETH.

4.4. Characterization of preferences

Between the valid sample (210 individuals): 100 (47.6%) preferred expressing value judgments in words, associated with MACBETH task; 46 (21.9%) preferred expressing value judgments numerically, associated with TTO task; and 64 (30.5%) were indifferent.

Regarding to the main comments obtained (total of 74), most of individuals considered MACBETH task easier (13 individuals) since it allows health states and problems comparison side by side and had difficulties in quantifying qualitative features related to health (17 individuals). However some individuals preferred the TTO exercise and considered this task easier (6 individuals) and more direct, not as subjective to individual's interpretation (6 individuals).

4.5. Influence of numeracy

A. On preferences

Individuals with lower numeracy level (0-1 right answers) are mostly indifferent between the two methods; individuals with higher numeracy level (2-3 right answers) mostly preferred the MACBETH task. To determine the influence of numeracy level in individuals' preferences we performed a multinomial logistic regression, with

the dependent variable preference (categorical variable with three levels: MACBETH; TTO and indifferent) and the independent variable numeracy level (numerical variable that varies between 0 and 3). Using as reference level the indifferent category, numeracy level allows for a statistically significant distinction between MACBETH and the reference level (OR = 2.50 (95% CI: 1.63-3.86)). That means that, individuals with higher numeracy level are more likely to prefer expressing value judgments in words than being indifferent. However, a distinction is not possible between reference level (indifferent) and preference for expressing value judgments in numbers (OR = 1.30 (95% CI: 0.87-2.14)). Additionally, using as reference level the TTO category, the numeracy level allows for a statistically significant distinction between MACBETH and the reference level (OR = 1.83 (95% CI: 1.13-3.00), so higher numeracy level increases the probability to prefer expressing value judgments in words, in comparison with numbers.

B. On health states evaluations

In general, few differences were obtained between both methods (TTO and MACBETH) for the two subsamples: low numeracy sample (0-1 right answers) and high numeracy sample (2-3 right answers). The MAD for TTO is 0.120 and the MAD for MACBETH is 0.128.

According to Table 4, individuals with higher numeracy levels showed high correlation coefficients between the MACBETH and TTO scales, however in general these correlation coefficients are high for the whole sample (with exception for subsample "0 right answers" that is very small).

Table 4- Spearman's correlation coefficients between MACBETH and TTO mean values scales, according to responses obtained in the numeracy test.

	ρ (valor p)	**ρ^2**
0 right answers	0.378 (0.11)	0.143
1 right answer	0.905 (2.01e-06)	0.819
2 right answers	0.895 (2.14e-06)	0.801
3 right answers	0.965 (8.98e-07)	0.931

C. On the exclusion of answers

In general, the number of exclusions identified for a low numeracy sample and for a high numeracy sample is not statistically significant (at a 5% level). However, an interesting result is that in terms of the additional criteria analyzed in relation to TTO it was found that low numeracy sample have a higher number of logical inconsistencies (60% vs. 40%).

5. Discussion

The main focus of this study was to develop a new non-numerical protocol for valuing health states, based on the MACBETH approach; and analyze congruence across MACBETH approach and a conventional method, the TTO. From the comparison between the two methods in terms of exclusions identified and dropouts, mean values scales and preferences we concluded that: although, in first analysis, MACBETH revealed a higher number of exclusions, in fact exclusion criteria defined differed across the two methods. If we consider the additional criteria "logical inconsistency" analyzed for TTO that are similar to the logical inconsistency considered in MACBETH and that leads to 21 exclusions (8.6%) we identified 126 responses (51.9%) with logical inconsistencies in TTO exercise and from those 10 (4.1%) are serious. This is a higher number than that obtained by MACBETH. We also noted that the most of exclusion associated with MACBETH resulted from inconsistencies in the matrix of judgments, in relation to that a suggestion is made to include in the MACBETH survey protocol, simultaneously with judgments elicitation, a check for inconsistencies so individuals can correct them immediately.

Analyzing the dropouts we realized that if the MACBETH is the first displayed task a higher number of individuals completed it, which may suggest ease of understanding and response to the MACBETH task. This conclusion is enforced by the most preference obtained in relation to this method and by the written comments collected.

In relation to the two mean value scales, a strong correlation coefficient was obtained. In general, mean values scales are consistent with each other; however, MACBETH mean values are in general higher, particularly, for health states with more limitations (severe health states). For health state 33333 the value obtained by TTO is -0.446 and for MACBETH is -0.080. This difference may also due to rescaling of the negative values originally obtained through the monotonic transformation, since the original scales vary between different points.

Despite the high correlation between the two mean values scales, at individual level, the MACBETH proposed scale is not consistent with the corresponding TTO scale.

Other point of this study was to compare the mean value scales obtained with the Portuguese TTO tariffs for EQ-5D previous elicited in [22], in order to check the consistency of the obtained values. We calculated the Spearman rank order correlation coefficient (Table 5) to get an overall picture of the consistency of ranking across results and it appears that there is a high degree of correlation – all coefficients being significant.

Table 5- Pearson's correlation coefficients between mean value scales obtained in this study and the Portuguese tariff.

	TTO *vs.* TTO PT	**MACBETH *vs.* TTO PT**
r	0.92	0.96
p-value	6.4e-11	1.8e-14

Regarding to the study of the influence of the numeracy level we concluded that, in general, both individuals with higher and lower numeracy levels preferred expressing value judgments in words; and with an increasing level of numeracy that preference also increases, in relation to expressing value judgments in numbers and being indifferent. Considering a previous study [16] it should be expected that individuals with a higher numeracy level preferred expressing values in numbers and individuals with high fluency level preferred expressing judgments in words. To approach more this point an additional fluency test is recommended for future research.

6. Conclusions

Results suggest that it is worth considering the use of non-numerical preference elicitation methods in health, highlighting the fact that obtained values are consistent and individuals shown higher preference for this mode of expressing value judgments.

Besides that, these methods are cognitively less demanding and present a simplified and unique protocol for states better and worse than death.

Further research may be directed to improvement of the non-numerical MACBETH protocol, to replicate the method in a larger sample of the Portuguese population and understand eventual advantages of non-numerical methods in special disease populations, diseases states and across population from different countries.

References

[1] D. Banta, "The development of health technology assessment," *Health Policy (New. York).*, vol. 63, pp. 121–132, 2003.

[2] T. Walley, "Health technology assessment in England: assessment and appraisal," *MJA*, vol. 187, no. 5, pp. 283–285, 2007.

[3] K. Stein, A. Fry, A. Round, R. Milne, and J. Brazier, "What Value Health? A Review of Health State Values Used in Early Technology Assessments for NICE," *Appl. Health Econ. Health Policy*, vol. 4, no. 4, pp. 219–228, 2005.

[4] S. J. Whitehead and S. Ali, "Health outcomes in economic evaluation: the QALY and utilities.," *Br. Med. Bull.*, vol. 96, pp. 5–21, Jan. 2010.

[5] M. C. Weinstein, G. Torrance, and A. McGuire, "QALYs: the basics.," *Value Health*, vol. 12 Suppl 1, pp. S5–9, Mar. 2009.

[6] M. J. Drummond, Michael F. Sculpher, G. W. Torrance, B. J. O'Brien, and G. L. Stoddart, *Methods for the Economic Evaluation of Health Care Programmes*, Third Edit. Oxford University Press, 2005.

[7] D. L. Patrick, H. E. Starks, K. C. Cain, R. F. Uhlmann, and R. a. Pearlman, "Measuring Preferences for Health States Worse than Death," *Med. Decis. Mak.*, vol. 14, no. 1, pp. 9–18, Feb. 1994.

[8] P. Dolan, C. Gudex, P. Kind, and a. Williams, "Valuing health states: A comparison of methods," *J. Health Econ.*, vol. 15, no. 2, pp. 209–231, 1996.

[9] P. Krabbe, M.-L. Essink-Bot, and G. Bonsel, "The Comparability and Reliability of Five Health-State Valuation Methods," *Soc. Sci. Med.*, vol. 45, no. 11, pp. 1641–1652, 1997.

[10] G. W. Torrance, "Preferences for health outcomes and cost-utility analysis.," *The American journal of managed care*, vol. 3 Suppl. pp. S8–20, May-1997.

[11] K. Tolley, "What are health utilities?," *Hayward Med. Commun.*, no. 4, pp. 1–8, 2009.

[12] G. W. Torrance, W. Furlong, and D. Feeny, "Health utility estimation.," *Expert Rev. Pharmacoecon. Outcomes Res.*, vol. 2, no. 2, pp. 99–108, Apr. 2002.

[13] C. Green, J. Brazier, and M. Deverill, "Valuing health-related quality of life. A review of health state valuation techniques.," *Pharmacoeconomics*, vol. 17, no. 2, pp. 151–65, Feb. 2000.

[14] S. Ali and S. Ronaldson, "Ordinal preference elicitation methods in health economics and health services research: using discrete choice experiments and ranking methods.," *Br. Med. Bull.*, vol. 103, no. 1, pp. 21–44, Sep. 2012.

[15] L. Rudmik and M. Drummond, "Health economic evaluation: important principles and methodology," *Laryngoscope*, vol. 123, no. 6, pp. 1341–7, Jun. 2013.

[16] B. Fasolo and C. a. Bana e Costa, "Tailoring value elicitation to decision makers' numeracy and fluency: Expressing value judgments in numbers or words," *Omega*, vol. 44, pp. 83–90, Apr. 2014.

[17] S. Woloshin, L. M. Schwartz, M. Moncur, S. Gabriel, and a. N. a. Tosteson, "Assessing Values for Health: Numeracy Matters," *Med. Decis. Mak.*, vol. 21, no. 5, pp. 382–390, Oct. 2001.

[18] S. R. Schwartz, J. McDowell, and B. Yueh, "Numeracy and the shortcomings of utility assessment in head and neck cancer patients.," *Head Neck*, vol. 26, no. 5, pp. 401–7, May 2004.

[19] C. a. Bana E Costa, J.-M. De Corte, and J.-C. Vansnick, "Macbeth," *Int. J. Inf. Technol. Decis. Mak.*, vol. 11, no. 02, pp. 359–387, Mar. 2012.

[20] C. Bana e Costa, J.-M. De Corte, and J. Vansnick, "M-MACBETH - Guia do utilizador," pp. 1–56, 2005.

[21] N. Bansback, A. Tsuchiya, J. Brazier, and A. Anis, "Canadian Valuation of EQ-5D Health States: Preliminary Value Set and Considerations for Future Valuation Studies.," *PLoS One*, vol. 7, no. 2, Jan. 2012.

[22] L. N. Ferreira, P. L. Ferreira, L. N. Pereira, and M. Oppe, "The valuation of the EQ-5D in Portugal.," *Qual. Life Res.*, vol. 23, no. 2, pp. 413–23, Mar. 2014.

[23] R. Brooks, "EuroQol: the current state of play.," *Health Policy*, vol. 37, no. 1, pp. 53–72, Jul. 1996.

[24] P. Dolan, "Modelling valuations for EuroQol health states," *Med. Care*, vol. 35, no. 11, pp. 1095–1108, 1997.

[25] L. N. Ferreira, P. L. Ferreira, L. N. Pereira, and M. Oppe, "EQ-5D Portuguese population norms.," *Qual. Life Res.*, vol. 23, no. 2, pp. 425–30, Mar. 2014.