

# The social value of health programs: Is age a relevant factor?\*

Eva Rodríguez<sup>†</sup> and José Luis Pinto<sup>‡</sup>

University of Vigo

Pompeu Fabra University

---

\* The authors would like to thank Han Bleichrodt, Gregorio Serrano, Pedro Delicado and two anonymous referees for their comments and suggestions. Any remaining errors are the authors' responsibility. Jose Luis Pinto acknowledges financial support from Dirección General de Enseñanza Superior, SEC 98-0296-C04-01. Eva Rodríguez acknowledges financial support from University Complutense of Madrid in the form of a grant received while this study was conducted.

† Correspondence to: Departamento de Economía Aplicada, Facultad de Economía, Universidad de Vigo, Lagoas-Marcosende s/n, 36200 Vigo, Galicia, España. Tel: + 34 986 812 516; fax: + 34 986 812 401; E-mail: emiguez@uvigo.es

‡ Pompeu Fabra University

## Abstract

In cost-effectiveness analysis (CEA) it is usually assumed that a QALY is of equal value to everybody, irrespective of the patient's age. However, it is possible that society assigns different social values to a QALY according to who gets it. In this paper we discuss the possibility of weighting health benefits for age in CEA. We also examine the possibility that age-related preferences depend on the size of the health gain. An experiment was performed to test these hypotheses. The results assessing suggest that the patient's age is a relevant factor when assessing health gains.

KEY WORDS: QALY; age weights; Social Welfare Function.

JEL: I18, D61,D63

# 1 Introduction

A quality-adjusted life year (QALY) is an output measure used in economic evaluations of health care programs. An advantage of this measure is that it combines the value of the quality and quantity of life in a single index number. By calculating the cost per QALY, it is possible to compare different health care programs in terms of efficiency.

Some authors<sup>1-4</sup> claim that this methodology makes a particular distributional assumption when it is used to allocate social resources. The QALY approach assumes that a QALY has equal social value to everybody irrespective of the patient's characteristics (income, social class, age, etc.). In this paper, we analyse the extent to which one of these characteristics, i.e. the patient's age, might be considered a relevant factor when QALYs are socially evaluated.

There are two main lines of thinking concerning age-related preferences. One of them, is based on efficiency.<sup>5-7</sup> The other is based on equity.<sup>8</sup> Age weights based on efficiency reflect the relationship between the patient's age and his or her social role. Children and, frequently, older patients, need physical and financial support from others. Younger and middle-aged adults are those who support this burden and thereby contribute more to social

welfare. To take account of this, Murray and Lopez<sup>6</sup> suggest “*unequal age-weights as an attempt to capture different social roles at different ages*”.

According to Williams,<sup>8</sup> age weights are derived from an equity argument which “*reflects the feeling that everyone is entitled to some ‘normal’ span of health [in terms of both quality and quantity of life](...). The implication is that anyone failing to achieve this has in some sense been cheated, whilst anyone getting more than this is ‘living on borrowed time’*”. With this argument younger patients —with other things being equal— will have higher weights, because they have completed a smaller portion of this normal span.

Each argument has very different implications. Based on Murray’s reasoning, weights should be higher for younger and middle-aged adults and lower for children and older people. However, Williams’ reasoning implies monotonically decreasing weights with increasing age.

Regarding empirical findings, different studies have confirmed that people take the patient’s age into account when they are asked to evaluate social health programs —see Tsuchiya<sup>9</sup> for a review and discussion of existing empirical studies. However, these studies provide only limited evidence. Most of them did not obtain a quantifiable weight,<sup>10</sup> and those that did, did not consider the possibility that the weights might depend on the health gains

being evaluated. This possibility also warrants further investigation. For example, one may be indifferent between giving one year of life to two people of different ages each with a life expectancy of one year. However, if their life expectancy is 10 years, we may favour the younger.

There are several examples in the literature<sup>6,11-13</sup> of age weights which have been estimated without taking into account the potential effect of the size of the health gain. For example, Busschbach<sup>13</sup> found that for a health gain of 2 years the relative age weights of a patient 70 years old with regard to a patient 35 years old, was 0.7. It is uncertain, however, whether giving 10 years to a 70 year old is equivalent to giving 7 years to a 35 years old, that is whether the relative age weights remain constant.

Murray<sup>6</sup> also assumes that age-related preferences are independent of health gains, but in a different way. His age weights are derived from the function

$f(a) = C * a * \exp(-\beta * a)$  —where  $a$  is age and  $C$  and  $\beta$  are parameters— and by integration he obtains the social values assigned to a gain of  $t$  years for a person who is  $a$  years old. So, for  $C = 0.16$ ,  $\beta = 0.04$  and a 3% discount rate, giving 2 years to a 35 years old is equivalent to giving 4.25 years to a 70 years old. This is because the weight of the year going from 30 to 31 plus

the weight of the year going from 31 to 32 is equivalent to the sum of the weights of the years going from 70 to 71, 71 to 72 and so on. Likewise, it can be estimated that 5 years for a 35 year old is equivalent to 13 years for a 70 year old. In the first example, the relative social value of the health gain of a patient 70 years old with regard to a patient 35 years old is 0.47 ( $2/4.25$ ) and in the second example it is 0.38 ( $5/13$ ). The proportion is not fixed but this does not mean that age weights depend on the health gain. So, for example, the weight for the year from 70 to 71 is always 50% of the year from 35 to 36, independently of how many years are to follow.

This paper has several objectives. One is to develop a model that shows how the traditional QALY model should be modified to incorporate age weights. The second is to test the existence of age-related preferences. The third is to test whether the relative social value of different health gains for different ages is proportional. The fourth is to test whether the relative social value should be estimated using the assumption of constant age weights, as Murray does. Section two will be devoted to developing a theoretical model and to presenting the hypotheses to be tested. In section three, an experiment aimed at testing these hypotheses is described. In section four the results of the experiment are discussed and, finally, conclusions are presented.

## 2 QALY model

### 2.1 Preliminary remarks

Before formalising the aggregate QALY model, it might be useful to make some clarifications. As mentioned in the introduction, a QALY is a health output measure. However, given that its theoretical foundations come from Utility Theory,<sup>14–16</sup> it can also be considered a utility measure. Although at an individual level, this difference is not relevant, at a social level it becomes important. If a QALY is considered as a health output measure, the total output produced by a health care program can be obtained simply by adding the individual QALYs generated by the program. However, if a QALY is considered a utility measure, the problem of interpersonal comparisons of utilities arises.

Following Sen,<sup>17,18</sup> Bleichrodt<sup>2</sup> suggests that to incorporate equity considerations in the cost–utility analysis based on QALYs, utility has to be considered as a cardinal measure and be fully comparable. Therefore, the first area of concern is the extent to which a QALY is a cardinal measure. Torrance<sup>19</sup> identifies different methods to obtain QALYs that allow them to be interpreted as cardinal measures —*Standard Gamble* in an uncertain

world and *Time Trade-Off* in a certain world. To make interpersonal comparisons of utilities he proposes assigning a value of zero to death and a value of one to full health wilst considering that full health has the same value for everyone.

Having defined a QALY as a cardinal measure which is comparable between individuals, we can identify the assumptions which implicitly underlie the aggregation of QALYs in economic evaluation.

## 2.2 The aggregate QALY Model

The output of a health care program can be defined as a distribution of health gains, measured in QALYs, that are provided to a population. We assume a population of  $n \geq 3$  people and denote by  $T \in \mathfrak{R}_+^n$  the set of possible results from different health care programs. An element of  $T$  is defined by a vector,  $\tau = (t_1, \dots, t_n)$ , where  $t_i \in \mathfrak{R}_+$ ,  $i = 1, \dots, n$  indicates the number of QALYs received from a health care program by individual  $i$ .

The QALY model assumes the following value function to represent social preferences for a health care program,

$$V(\tau) = \sum_{i=1}^n t_i. \quad (1)$$



In order to introduce a temporal discount rate, it can be assumed that  $t_i$  refers to QALYs that have already been discounted.

It is then necessary to determine under what conditions social preferences can be represented by this value function. This can be done by considering a social planner (SP) whose preferences can be assumed to be a good approximation of society's preferences. Therefore, starting from the SP's preferences, a Social Welfare Function (SWF) can be obtained.

When deciding on his preferences for different health care programs, the SP takes into account the number of individual QALYs provided by each program. It is then necessary to obtain the preference relationship,  $\succeq$ , of the SP over elements in the set  $T$ , that is over  $\tau$ . This preference relationship is complete and transitive, that is to say  $\succeq$  is assumed to be a weak order. It can therefore be represented by a value function —SWF— denoted by  $V(\tau)$ . Our assumptions imply that  $\tau \succeq \tau'$  is equivalent to  $V(\tau) \geq V(\tau')$ . In addition, the SP's preferences depend positively on individual gains (QALYs). This consideration is formalised as follows:

**Assumption 1** Pareto assumption. The SP's value function is strictly increasing in  $t_i$ .

To obtain the additive function of the aggregate QALY model, an assump-

tion of independence must be made. Suppose that we divide the coordinates of vector  $\tau = t_1, \dots, t_n$  into two subvectors  $(t_{ij}, \bar{t}_{ij})$ , where  $t_{ij} = (t_i, t_j)$  and  $\bar{t}_{ij}$  is its complement.

**Definition 1** The vector  $t_{ij}$  is preferentially independent of its complement

$\bar{t}_{ij}$ , if and only if the preferences over  $t_{ij}$ , given  $\bar{t}_{ij}$ , do not depend on  $\bar{t}'_{ij}$ , for every  $\bar{t}'_{ij}$ . That is,

$$\left( t'_{ij}, \bar{t}'_{ij} \right) \succeq \left( t''_{ij}, \bar{t}'_{ij} \right) \Rightarrow \left( t'_{ij}, \bar{t}_{ij} \right) \succeq \left( t''_{ij}, \bar{t}_{ij} \right), \quad \text{all } \bar{t}_{ij}, \bar{t}'_{ij}, t'_{ij}, t''_{ij}.$$

It seems appropriate to consider that the SP's preferences over two different distributions of health, that differ only in the health gain of two individuals, depend only on the SP's preferences regarding the gains of those two patients. Formally stated,

**Assumption 2** The vector  $t_{ij}$  is preferentially independent of its comple-

ment  $\bar{t}_{ij}$  for all  $i \neq j$ .

These assumptions allow the postulation of an additive value function for the SP, using the following theorem:

**Theorem 1 (Debreu<sup>20</sup>)** The SP's value function,  $V$ , is additive,

$$V(\tau) = \sum_{i=1}^n v_i(t_i),$$

if and only if the SP's preference relationship verifies assumptions 1 and 2.

Where  $v_i$  is a positive monotonic transformation defined over  $t_i$  that reflects the interpersonal comparisons made by the SP.

It must be taken into account that in the QALY model defined in equation (1),  $v_i(t_i)$  is common to all patients and equal to  $t_i$ . Therefore, it is necessary to make some additional assumptions to obtain a complete characterisation of QALYs. To do that, the concept of permutation can be introduced.

Given  $\tau = (t_1, \dots, t_n)$ , a permutation of  $\tau$  denoted by  $\sigma$  is defined as an exchange of individual values in the set of patients,  $\{1, \dots, n\}$ . Then,  $\tau_\sigma = (t_{\sigma(1)}, \dots, t_{\sigma(n)})$  is a permutation, where  $t_{\sigma(i)}$  is the permuted value of patient  $i$ ,  $i = 1, \dots, n$ .

**Assumption 3.** Anonymity.  $\tau \sim \tau_\sigma$  for all health distributions  $\tau$  and for all permutations  $\sigma$ .

This assumption states that, if a distribution of health gains is a permutation of a given distribution of health gains, the SP must be indifferent to both. For instance, given two patients,  $i$  and  $j$ , the SP is indifferent to distribution  $(a, b)$  and  $(b, a)$ , where the first element refers to the QALYs received

by  $i$  and the second to those received by  $j$ .

Next, we introduce a scaling assumption. Given that  $v_i(t_i)$  is increasing in  $t_i$  and  $t_i \geq 0$ , then  $t_i = 0$  is the number of health gains least preferred by the SP. Therefore, we will consider that the social value of  $t_i = 0$  is equal to zero,  $v_i(0) = 0$  for all  $i$ .

**Proposition 1** Under assumptions 1, 2, 3 and scaling assumption, we have

that,  $v_i(t_i) = v_j(t_j) \equiv v(t)$ , for all  $i, j$  and  $t_i = t_j \equiv t$ .

The proof can be found in appendix A.

This proposition guarantees that a given health gain will have the same social value independently of who receives it. That is,

$$V(\tau) = \sum_{i=1}^n v(t_i).$$

Finally, the QALY model is reached if a final assumption is imposed, i.e. that an additional QALY always has the same social value. That is, only the total number of QALYs matters. Formally stated,

**Assumption 4.**  $v(t_i)$  can be represented by any linear function over  $t_i$ .

If we choose the function  $v(t_i) = t_i$ , the model QALY —equation (1)— is obtained.

Starting from this characterisation, we introduce some modifications that allow us to take into account the influence of patient age in the SP's preferences. The following subsection is devoted to this purpose.

### 2.3 The weighted QALY Model

Suppose that the SP's preferences —and, therefore, the SWF—, depend, not only on individual QALYs, but also on patient age. Under this assumption, we need to obtain the SP preference relation,  $\succeq$ , defined on the set  $A \times T$ . One element on this set, denoted as  $(\mathbf{a}, \tau)$ , can be obtained from an element  $\tau = (t_1, \dots, t_n)$  through  $((a_1, t_1), \dots, (a_n, t_n))$ , where  $\mathbf{a} = (a_1, \dots, a_n) \in \mathfrak{R}_+^n$  is the age vector.

Given that the SP preference relation over  $A \times T$  is complete, and transitive, it can be represented by means of a value function that we denote by  $W(\mathbf{a}, \tau)$  such that,  $(\mathbf{a}, \tau) \succeq (\mathbf{a}', \tau')$  is equivalent to  $W(\mathbf{a}, \tau) \geq W(\mathbf{a}', \tau')$ .

We maintain assumptions 1 and 2, but with slight modifications. Let us denote these assumptions as assumption 1' and assumption 2'. Now, assumption 1' states that  $W$  is strictly increasing in  $t_i$ , given  $\mathbf{a}$ . Assumption 2' states that the SP's preferences between two health distributions that are different only in terms of the health gains received by two patients, depend only on

the individual gains of both patients and on interpersonal comparisons of utility, based on age, made by the SP. That is, assumption 2 is maintained but now takes into account that  $t_{ij} = ((a_i, t_i), (a_j, t_j))$ .

Under these assumptions, Theorem 1 is verified and allows us to obtain the following additive value function for the SP:

$$W(\mathbf{a}, \tau) = \sum_{i=1}^n w_i(a_i, t_i),$$

where  $w_i$  is a positive monotonic transformation defined over  $t_i$  that reflects the interpersonal comparisons made by the SP, but takes into account patient age.

Now, we adopt an additional assumption. Given any distribution of QALYs,  $\tau = (t_1, \dots, t_n)$ , let  $\delta$  denote any permutation of individual values among patients of the same age in the set  $\{1, \dots, n\}$ , so that

**Assumption 5.** Conditioned anonymity.  $\tau \sim \tau_\delta$  for any distribution  $\tau$  and any permutation  $\delta$ .

This weakens the anonymity assumption that is implicitly present in the aggregate QALY model, defined in equation (1). Now, a distribution of QALYs, which is a permutation of other given distributions of QALYs, can

have a different social valuation depending on the patient's age. This causes a reconsideration of the scaling assumption.

Given an age,  $a_o$ , that we call the "reference" age,  $w_i(a_o, 0) = 0$  for all  $i$  such that  $a_i = a_o$ .

**Proposition 2** Under assumptions 1', 2', 5 and the scaling assumption,

$$w_i(a_i, t_i) = w_j(a_j, t_j) \equiv w(t), \text{ for all } a_i = a_j \equiv a_o \text{ and } t_i = t_j \equiv t.$$

The proof is identical to the proof of proposition 1, given that the only difference is that in proposition 2 we introduce a constant,  $a_o$ .

Given that the value function is the same for patients of the same age, it is possible to reduce the dimensionality of the problem, transforming the QALYs received by patients of different ages into QALYs received by patients in the reference age. Let us fix as a reference age the age of patient  $h$ ,  $a_o = a_h$ , and suppose that, for each  $(a_i, t_i)$ , there exists a value,  $t_{h(i)}^*$ , that makes the SP indifferent between allocations  $(a_i, t_i)$  and  $(a_o, t_{h(i)}^*)$ . Then  $w_i(a_i, t_i) = w_h(a_o, t_{h(i)}^*) \equiv w(t_{h(i)}^*)$ . Naturally  $a_o$  has to be suitably chosen such that a value of  $t_{h(i)}^*$  always can be found. In the experiment described in next section  $a_o = 20$ .

Therefore, starting from the aggregate QALY model, we have obtained

the following model, which we call the weighted QALY model,

$$W(\mathbf{a}, \tau) = \sum_{i=1}^n w_i(a_i, t_i) = \sum_{i=1}^n w(t_{h(i)}^*), \quad (2)$$

where  $t_{h(i)}^* = g(a_i, t_i; a_o)$ .

With this framework, a distribution of health gains amongst patients of different ages can be expressed as a distribution of health gains amongst patients of the same age.

The weighted QALY model, defined in equation (2), generalises the aggregate QALY model, equation (1), in the sense that the latter can be obtained as a particular case of the former when age weights are equal for everyone,  $t_{h(i)}^* = t_i$ , and assumption 4 is maintained.

## 2.4 Hypotheses about age-related preferences

With the help of this theoretical model we will present hypotheses about age-related preferences and the tests we need to conduct. To avoid unnecessary notational complication we will omit the sub-index  $i$  and  $h(i)$  from now on.

**Hypothesis 1. Existence of age-related preferences.** To test for the existence of age weights we calculate  $t^*$  for different ages and test if  $t^* = t$ .



**Hypothesis 2. Constancy of the age-related relative social value of health gains.** To test if the relative social value of different health gains for different ages is constant we calculate  $t^*/t$ , and estimate if, given  $a$ , it is constant for all  $t$ .

**Hypothesis 3. Independence of age-related preferences from the size of health gains.** To test if this hypothesis holds we define  $g(a, t; a_o)$  as  $g(a, t_s; a_o)$  when  $t$  is sufficiently small. We then say that age-related preferences are independent of the size of the gains if

$$g(a, t; a_o) = \int_{a=a}^{a+t} g(a, t_s; a_o) da. \quad (3)$$

Murray's function,  $f(a)$ , is similar to the function  $g(a, t_s; a_o)$ . Eq. (3) says that the relative social value of a health gain  $t$  for age  $a$  can be estimated as the sum of successive values of small health gains from  $a$  to  $a + t$ .

## 3 The experiment

### 3.1 Design

To test our hypotheses, we conducted an experiment with 61 undergraduate students—21 from the Economics department, 20 from the Political Science department and 20 from the Law school. The students were paid approximately \$18 for participating. The experiment consisted of three meetings with participants, each meeting on a different day. During the first meeting the objective of the experiment was explained to the participants. They, then answered a pilot questionnaire to familiarise them with the type of questions they would have to answer in the subsequent meetings.

The second meeting was organised in different sessions, each one with five participants. We showed each participant a list with eleven different health care programs. Each program consisted of a different pair  $(a, t)$ , where  $a$  is the patient's age, and  $t$  is the health gain, measured in years in full health, provided to this patient. Health gains were measured in years in full health in order to make the task easier for respondents. For each program, each participant had to decide the increment of years,  $t^*$ , that would leave him or her indifferent between the proposed program  $(a, t)$  and the program  $(20, t^*)$ ,

where  $a = 20$  is the reference age. This is known as the matching technique.

In the pilot questionnaire we detected that the participants had some difficulties in choosing a specific number  $t^*$ , so we decided to use “choice–bracketing”. This consists of approaching the value through successive questions where it is always necessary to choose between two allocations —see appendix B. Sometimes the balancing mechanism does not allow the elicitation of an exact value for  $t^*$ , but produces an interval instead. In this case, the mid–point value of the interval is taken.

The construction of a list of health care programs for evaluation was not easy because there is a trade–off between the number of programs to be evaluated and the participants’ degree of concentration. To avoid this problem we selected four ages that we considered representative of different periods of human life: 1, 20, 40 and 60 years old; and four health gains —measured as healthy life–years—: 2, 10, 20 and 40 years. Given that the age of 20 years was used as a reference age and the pair  $(a, t) = (60, 40)$  is unrealistic, we have the following eleven allocations —health care programs— for evaluation:  $(1, 40)$ ,  $(1, 20)$ ,  $(1, 10)$ ,  $(1, 2)$ ,  $(40, 40)$ ,  $(40, 20)$ ,  $(40, 10)$ ,  $(40, 2)$ ,  $(60, 20)$ ,  $(60, 10)$ ,  $(60, 2)$ . The first element refers to the patient’s age and the second refers to years in full health given to this patient.

In any experiment of this kind it is important to analyse the extent to which the use of another technique provides similar results —consistency across methods. For that reason, after applying the choice–bracketing to all programs, we provided each participant with fifteen cards that they had to rank from more to less preferred —direct ordering technique (DO). Each card corresponded to each of the previously assessed programs. Note that now there are four more cards than programs in the previous list. These four additional cards correspond to programs that provide the selected health gains to 20–year–old patients. In addition, each participant is asked to justify his ranking with a short written explanation.

Two weeks later, we organised a third meeting in order to check test–retest reliability. All participants had to repeat some of the tasks from the previous meeting. Participants were divided into three groups of 20 —one of the previous participants did not return. Each group matched four of the eleven programs —a program was evaluated by two groups— and ranked five of fifteen possible cards. Both the cards to be evaluated and those to be ranked were chosen randomly for each group.

## 3.2 Method of analysis

### Testing the hypotheses

To test for the existence of age-related preferences we estimated the average  $t^*$ , denoted as  $\bar{t}^*$ , for each of the eleven combinations. If  $\bar{t}^* = t$  we rejected the hypothesis of the existence of implied age weights. A Student t-test was used to compare the two values.

To test whether the relative social value of different health gains for different ages was constant we calculated  $\bar{t}^* / t$  for each of the four durations and for each age group. If  $\bar{t}^* / t$  was not constant in each age group we rejected the hypothesis of constant relative social value. We used an F test for each age group to test for constancy.

Testing the assumption of independence of age-related preferences from the size of health gains involved three steps.

1. We estimate  $g(a, t_s; a_o)$  for  $t_s = 2$  years using the answers,  $t^*$ , from the 2 year gain, here denoted as  $\hat{g}_2(a; a_o)$ . Of course, based on equation (3) infinitesimal health gains should be used to produce  $g(a, t_s; a_o)$ . However, this is not possible as people do not have the capacity to discriminate to this degree, and it is necessary to use small health gains that can be used in matching questions. A period of 2 years was chosen as a sufficiently small

health gain and we assumed that relative age weights would remain the same for smaller health gains. To estimate  $\hat{g}_2(a; a_o)$  Ordinary Least Squares (OLS) regression was used.

2. The value of  $g(a, t; a_o)$  for the 15 combinations was predicted as follows,

$$\hat{t}_p^* = \hat{g}(a, t; a_o) = \sum_{i=0}^{(t/2)-1} \hat{g}_2(a + 2i; a_o).$$

3. Predicted social values,  $\hat{t}_p^*$ , were compared with the real ones obtained from the survey,  $\bar{t}^*$ , to determine whether they were statistically equal. We used the  $T$  statistic to conduct this test which follows a  $\chi^2$  distribution.

### **Testing the consistency of the responses**

Responses were considered consistent when:

1. The ranking of cards obtained by direct ordering and the implicit ranking obtained from the matching exercise coincided at the individual level. Spearman rank correlation coefficient (SCC) was used for each participant and the average SCC of all participants was calculated.

2. The ranking of the cards obtained by direct ordering and the implicit ranking obtained from the matching exercise coincided at the aggregate level. Both orderings were aggregated using the Borda rule. The aggregate ordering obtained from the direct ranking will be denoted by S-DO and the aggregate

ordering obtained from the implicit ranking obtained from matching will be denoted as S-matching. To assess the degree of coincidence between both orderings, we used both SCC and the Kendall rank correlation coefficient (KCC).

3. Test-retest reliability is high. SCC for the direct ranking and the Pearson linear correlation coefficient (PCC) for the matching questions were used for each participant, and the average SCC and PCC of all participants were calculated.

### 3.3 Results

Table 1 shows that  $\bar{t}^* \neq t$  in 8 out of 11 cases, thereby confirming the existence of age weighting in our sample.

The ratios  $\bar{t}^* / t$ , can be estimated from Table 1 and are shown in Figure 1. The hypothesis of constancy of proportions is rejected ( $F=7.39$ ) at the 1% level in the case of the 1 year old, and in the case of the 40 years old ( $F=3.58$ ) at the 5% level but it is not rejected for the 60 years old ( $F=1.7$ ). In general then, the hypothesis is rejected. It can be seen in Figure 1 that these proportions follow a clear pattern, increasing for babies and decreasing for the elderly. It can also be seen that health gains for 20 and 40 year olds

are rated similarly for short duration health gains but receive a lower value in 40 year old for longer duration health gain. These results are consistent with the existence of an age weighting function similar to the one proposed by Murray.

To test for the assumption of independence of age-related preferences from health gains we:

1.-Estimate  $\widehat{g}_2(\cdot)$ . The function which displays the best goodness of fit was  $\widehat{g}_2(a; a_o) = \widehat{\alpha}_1 a^{\widehat{\alpha}_2} \exp(\widehat{\alpha}_3 a^2)$ . The estimated equation is:

$$\widehat{\ln g_2(\cdot)} = - \underset{(-4.934)}{0.3348} + \underset{(11.546)}{0.3708} \ln a - \underset{(-8.120)}{0.0003} a^2 \quad R^2 = 0.43$$

(student's  $t$  in brackets)

The function  $\widehat{g}_2(\cdot)$  is similar to the function  $f(\cdot)$  used by Murray, which has a inverted  $u$  shape and reaches a peak at age 25 years.

2.-Obtain a prediction of the value of  $g(a, t; a_o)$  for the 15 combinations evaluated. The vector integrated by these values is denoted as  $\widehat{t}_p^*$ .

3.- Finally, we tested whether the  $\widehat{t}_p^*$  vector and the vector of social values obtained from the sample,  $\bar{t}^*$ , can be considered statistically equal—see appendix C for a more detailed explanation of the statistical method used. The statistic used,  $T \sim \chi_{11}^2$ , received a value of 15.76, so the null hypothesis that the two sets of vectors are equivalent is accepted at the 5% level. For the



combinations  $(20, 2)$ ,  $(20, 10)$ ,  $(20, 20)$  and  $(20, 40)$  the confidence intervals  $(1.93 \pm 0.21; 9.74 \pm 1.04; 19.07 \pm 1.90; 33.77 \pm 2.80; 5\% \text{ level})$  contained the real value except in one case, reinforcing our conclusion that both sets of vectors are similar.

Table 2 shows high correlation coefficients, suggesting a high consistency across methods, both at the social and individual level. Test-retest reliability was also high (SCC= 0.89 and PCC=0.93).

## 4 Discussion

The results of the experiment reported show that respondents consider patient's age as a relevant factor in the valuation of health gains. This result leads us to reject the QALY model as defined in equation (1) in favour of the weighted QALY model —equation (2).

We also reject the hypothesis that the relative social value of health gains is constant when comparing people of different ages, making it necessary to estimate an age weighting function that has the effect of producing a change in the relative social value of health gains when these health gains change.

The method used by Murray allows us to estimate the relative value of

health gains in such a way. By method we mean the idea that age weights can—for large durations— be derived by integrating the implied age-weights for small durations. However, it assumes that implied age weights do not change with health gain. This is by no means a self evident assumption and for that reason we tested it explicitly. The test showed that the assumption that age-related preferences do not depend on the size of the health gain is a reasonable assumption. The fact that, in this sample, we did not find that the relative social values of different health gains were different from the values obtained from the age weighting function  $\hat{g}_2(\cdot)$  reinforces the empirical validity Murray’s method.

However, although Murray’s method can be adequate for generating age weights, it should be noted his age weights are purely based on efficiency reasons while in our experiment both efficiency and equity could be influencing the age-related preferences that we elicited.

The introduction of age weights in cost-effectiveness analysis (CEA) should not imply a major change in the usual methodology. In fact, age weights are already used, it is simply the case that the weight assigned is always one, so  $g(a, t_s; a_o)$  in the equation (3) is simply a horizontal line. It does not seem too complicated to weight the number of QALYs gained by each patient us-

ing a different  $g(a, t_s; a_o)$  function. It can be argued that the introduction of age weights may penalise young people when evaluating a medical treatment that affects people of different ages. For example, treatment for Parkinson's disease primarily affects elderly people, but some young people also have the condition. If treatment for Parkinson's disease is directed primarily at elderly patients, are we not penalizing younger people with the disease if we apply age weights? Again, this is not a problem caused by age weights because it is already present in CEA and is caused by the need to aggregate health outcomes from people in which treatment effectiveness varies. So if a group of the population is evaluated together with elderly people, they are going to be penalized if the final decision is taken using averages. However, it is clear that the introduction of age weights (different from one) may exacerbate this problem. In this case, the only way of not penalizing young people suffering from Parkinson's disease is to estimate CE ratios by age group. This is not unusual in CEA, given that cost/effectiveness ratios are usually estimated for subgroups of patients where costs and effectiveness differ.

It is important to stress the pilot nature of this experiment and therefore its limitations. In the present study a convenience sample was used, and the study needs to be repeated in a more representative sample before claiming

that the results adequately reflect social preferences. One of the main problems with our sample may be that the average age of the participants was 20. It is possible that age-related preferences would be influenced by respondent's age. However, previous studies that have addressed this topic<sup>12,13,23</sup> did not find significant differences in preferences. For example, Johannesson and Johannesson concluded that “[...] the trade-off between lives saved at different ages is unaffected by the age of the respondent.” Nevertheless, further research is clearly required to test for a possible influence of respondent age on age weights. A final limitation is that the function  $\widehat{g}_2(\cdot)$  was estimated by considering only the values that the participants assigned to four different ages and the remainder were fitted by the OLS model. It would be worthwhile increasing the number of ages, to get more precise estimates.

## 5 Conclusion

In this paper, we showed that the aggregate QALY model, which is commonly used in empirical studies, can be derived from some customary assumptions in the social choice literature. We also showed that replacing the assumption of anonymity by conditional anonymity introduces the possibility that health

gains can be weighted by age.

In the sample studied, the social value of equivalent health gains was considered to be different for people at different ages. Our results also show that the hypothesis that age-related preferences are independent of the size of the health gain is not rejected.

## Appendix A

Let  $h$  and  $p$  denote any two patients and  $t_h, t_p$ , the health gains received by each one from a health care program. Then, from assumptions 1, 2 and 3,

$$v_h(t_h) + v_p(t_p) + \bar{v}_{hp} = v_h(t_p) + v_p(t_h) + \bar{v}_{hp} \quad \text{for any } t_h, t_p$$

$$\text{where, } \bar{v}_{hp} = \sum_{i=1}^n v_i(t_i) - v_h(t_h) - v_p(t_p).$$

Given that  $(v_h - v_p)(t_h) = (v_h - v_p)(t_p)$  for any  $\{t_h, t_p\}$ , then,

$$(v_h - v_p)(t) = k \quad \text{for all } t, \text{ where } k \text{ is a constant.}$$

From condition 1,  $(v_h - v_p)(0) = 0$ . Starting from the last two equalities we obtain  $v_h(t) = v_p(t)$ , for all  $t$ .

## Appendix B

Part of the questionnaire we used can be found below. One of the eleven patient age and health gain combinations that the participants have assessed through the balance mechanism is included as an example.

In this section we will always show 2 treatments: A and B. The treatments are different from each other in the number of healthy life-years gained by the patient, and in the patient's age who receives the gains. You must say whether you prefer treatment A, treatment B, or are indifferent to both. Depending on your choice the questionnaire continues in the following way:

- If you choose an option which includes the word "stop", circle the word and go on to the next table (in which treatment A has been varied).

- If you choose an option which includes the word "continue", go on to the next line.

By way of simplification we will use the following notation:

Patient age= "Age"

Healthy life-year increases for the patient = "Years"

I prefer treatment A = "Pref. A"

I am indifferent to A and B = "Same"

I prefer treatment B = “Pref. B”

The treatments are the following:

<i>Treatment A</i>	<i>Treatment B</i>
--------------------	--------------------

<i>Age</i>	<i>Years</i>	<i>Age</i>	<i>Years</i>	Pref. A	Same	Pref. B
40	20	20	5	continue	stop	stop
40	20	20	40	stop	stop	continue
40	20	20	10	continue	stop	stop
40	20	20	30	stop	stop	continue
40	20	20	15	continue	stop	stop
40	20	20	35	stop	stop	continue
40	20	20	20	stop	stop	stop

## Appendix C

To test hypothesis 3 we need to estimate a different statistic to the more usual F, given that we are comparing two mean vectors that do not come from different samples. One is the mean vector originating from the sample interviews, whilst the other is the mean vector given by  $\hat{g}_2(\cdot)$ . Variability is

provided in one case by the variability of the answers and in the other case by the variability of the estimated parameters. For this reason it is necessary to use a statistic which permits comparison of these two types of means.

The T statistic was obtained as follows:

1.- The  $t_p^*$  function was defined as:

$$t_p^* : R^3 \longrightarrow \begin{matrix} R^{15} \\ t_{p1}^* (\alpha_1, \alpha_2, \alpha_3) \\ t_{p2}^* (\alpha_1, \alpha_2, \alpha_3) \\ \dots \\ t_{p15}^* (\alpha_1, \alpha_2, \alpha_3) \end{matrix}$$

2.- Assuming that

$$\sqrt{n} \left[ \begin{pmatrix} \widehat{\alpha}_1 \\ \widehat{\alpha}_2 \\ \widehat{\alpha}_3 \end{pmatrix} - \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{pmatrix} \right] \xrightarrow{d} N(0, V_{\widehat{\alpha}_1, \widehat{\alpha}_2, \widehat{\alpha}_3})$$

By the Cramer  $\delta$  Theorem<sup>21</sup>

$$\sqrt{n} (\widehat{t}_p^* - t_p^*) \xrightarrow{d} N \left( 0, (\nabla t_p^*)^t V_{\widehat{\alpha}_1, \widehat{\alpha}_2, \widehat{\alpha}_3} (\nabla t_p^*) \right)$$

where

$$\nabla t_p^* = \begin{bmatrix} \partial t_{p1}^* / \partial \alpha_1 & \dots & \partial t_{p15}^* / \partial \alpha_1 \\ \vdots & \dots & \vdots \\ \partial t_{p1}^* / \partial \alpha_3 & \dots & \partial t_{p15}^* / \partial \alpha_3 \end{bmatrix}$$



Then

$$\widehat{t}_p^* \xrightarrow{a} N \left( t_p^*, \underbrace{\left( (\nabla t_p^*)' V_{\widehat{\alpha}_1, \widehat{\alpha}_2, \widehat{\alpha}_3} (\nabla t_p^*) \right)}_{V_A} * 1/n \right)$$

3.- To test the null hypothesis,  $H_0 : \widehat{t}_p^* = \bar{t}^*$ , against the alternative,  $H_1 : \widehat{t}_p^* \neq \bar{t}^*$ , we used a t-test based on Mahalanobis' distance (Hausman test<sup>22</sup>) of the  $(\widehat{t}_p^* - \bar{t}^*)$  vector to one vector of zeros ( $15 \times 1$ ). The statistic used was

$$T = \left( \widehat{t}_p^* - \bar{t}^* \right)' (V_A + V_B)^{-1} \left( \widehat{t}_p^* - \bar{t}^* \right) \sim \chi_{15}^2$$

where  $V_A$  and  $V_B$  are the variance-covariance matrices estimated for  $\widehat{t}_p^*$  and  $\bar{t}^*$  respectively. For the combinations (20, 2), (20, 10), (20, 20) and (20, 40) there is no variance of real values. This generates numerical problems for estimating  $V_A + V_B$ , so that these combinations were excluded and  $T \sim \chi_{11}^2$ . In these four cases a check was made to ensure that the intervals predicted by function  $\hat{g}_2(\cdot)$  contained the real values.

## References

- [1] Ubel, P., Loewenstein, G., Scanlon, D. and Kamlet, M. Individual utilities are inconsistent with rationing choices: a partial explication of why Oregon's cost-effectiveness list failed. *Medical Decision Making* 1996; **16**:108–116.
- [2] Bleichrodt, H. Health utility indices and equity considerations. *Journal of Health Economics* 1997; **16**(1): 65–91.
- [3] Olsen, J.A. Persons vs years: two Ways of eliciting implicit Weights. *Health Economics* 1994; **3**: 39–46.
- [4] Loomes, G. and McKenzie, L. The use of QALYs in health care decision making. *Social Science Medicine* 1989; **28**(4): 299–308.
- [5] World Bank. *World Development Report 1993: Investing in Health*. Oxford University Press, 1993.
- [6] Murray, C. J. L. and Lopez, A. D. eds. *Global Comparative Assessments in the Health Sector: Disease Burden, Expenditures and Intervention Packages*, WHO, 1994.

- [7] Murray, C. J. L. and Acharya, A.K. Understanding DALYs. *Journal of Health Economics* 1997; **16**: 703–730.
- [8] Williams, A. Intergenerational equity: an exploration of the ‘fair innings’ argument. *Health Economics* 1997; **6**: 117–132.
- [9] Tsuchiya, A. Age-related preferences and age weighting health benefits. *Social science and Medicine* 1999;**48**(2): 267-276.
- [10] Nord, E., Richardson, J., Street, A., Kuhse, H. and Singer, P. Maximizing health benefits versus egalitarianism: An Australian survey of health issues. *Social Science and Medicine* 1995; **41**(10): 1429–1437.
- [11] Nord, E., Richardson, J., Street, A., Kuhse, H. and Singer, P. The significance of age and duration of effect in social evaluation of health care. *Health Care Analysis* 1996; **4**: 103–111.
- [12] Johannesson, M. and Johannsson, P-O. Is the valuation of a QALY gained independent of age? *Journal of Health Economics* 1997; **16**: 589—599.
- [13] Busschbach, J.J.V., Helsing, D. J. and Charro, F. T. The utility of health at different stages in life: A quantitative approach. *Social Science and Medicine* 1993; **37**(2): 153–158.

- [14] Torrance, G. W. Toward a utility theory foundation for health status index models. *Health Services Research* 1976; winter, 349–369.
- [15] PlisKin, J. S., Shepard, D. S. and Weintein, M. C. Utility functions for life years and health status. *Operations research* 1980; **28**: 206–24.
- [16] Torrance, G. W. and Feeny D. Utilities and quality-adjusted life years. *International Journal of Technology Assessment in Health Care* 1989; **5**: 559–575.
- [17] Sen, A. K. *Collective choice and social welfare*. Holden-Day, San Francisco, 1970.
- [18] Sen, A. K. On weights and measures: Informational constraints in social welfare analysis. *Econometrica* 1977; **45**: 1539–1572.
- [19] Torrance, G. W. Measurement of health state utilities for economic appraisal: a review. *Journal of Health Economic* 1986; **5**: 1–30.
- [20] Debreu, G. Topological methods in cardinal utility theory. In: K. J. Arrow, S. Karlin and P. Suppes, eds. *Mathematical methods in the social sciences*. Stanford University Press, 1960; 16–26.
- [21] Arnold, S.F. *Mathematical statistics*. Prentice-Hall, 1990.

- [22] Hausman, J. Specification tests in econometrics. *Econometrica* 1978; **46**: 1251–1271.
- [23] Cropper, M.L., Aydede, S. K. and Portney, P.R. Preferences for life saving programs: How the public discounts time and age. *Journal of Risk and Uncertainty* 1994; **8**: 243-265.

Table 1: Average social value of health gains (years) based on age

health gain, $t$	patient age, $a$	equivalent health gain for 20-yr-olds $\bar{t}^*$ (t-student <sup>(1)</sup> )
2	1	0.9 (-8.41)
	40	1.9 (-0.41)
	60	1.4 (-2.93)
10	1	5.4 (-7.45)
	40	10.1 (0.20)
	60	6.3 (-8.23)
20	1	13.2 (-5.55)
	40	18.6 (-1.03)
	60	10.9 (-10.58)
40	1	33.4 (-3.19)
	40	29.1 (-6.55)

(1)  $H_0: \bar{t}^* = t$ ;  $H_1: \bar{t}^* \neq t$   
 $n=61$

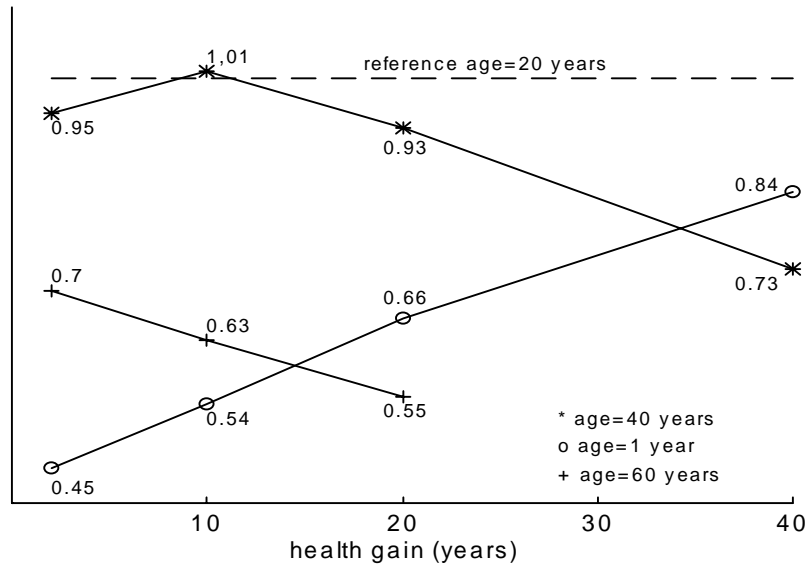


Figure 1: Weights of health gains based on age

Table 2: Ranking of health care programs<sup>(1)</sup>

<u><i>S – Matching</i></u> patient age , health gain (yrs)	<u><i>S – DO</i></u> patient age , health gain (yrs)
20,40	20,40
1,40	40,40
40,40	1,40
20,20	20,20
40,20	40,20
1,20	20,10
20,10	1,20
60,20	40,10
40,10	60,20
60,10	60,10
1,10	1,10
20,2	20,2
40,2	40,2
60,2	60,2
1,2	1,2

(1) From more to less preferred (average). n=61  
KCC=0.94 ; SCC=0.99 ; Individual SCC (average)=0.79