Social value of health programs: Is the age a relevant factor?*

Eva Rodríguez† and José Luís Pinto‡

December 1998

Abstract

A QALY —quality–adjusted life year— is an output measure used in economic evaluations of health care programs. The QALY approach assumes that a QALY has equal social value to everybody irrespective of the patient's age. However, it is possible that the general public assigns different social values to a QALY according to who gets it, derived from intergenerational equity judgements. In this paper we discuss the possibility of weighting health benefits by age in the QALY approach. In addition, we display the results of an experiment whose objective is to derive the age weights. The results of the experiment performed suggest that the patient's age is a relevant factor when the health gains are assessed. Moreover age weights are not constant, but they vary depending on gains considered.

*Preliminary version

†University of Vigo

‡Pompeu Fabra University

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 incorporates the value of quality and quantity of life into a single index number. Calculating the cost per QALY, it is possible to compare different health care programs in terms of efficiency.

Some authors (Ubel, Loewenstein et al, 1996; Bleichrodt, 1996; Olsen, 1994; Loomes and McKenzie, 1989; Wagstaff, 1991;...) claim that using this methodology to allocate social resources, provides some negative implications in terms of equity. The QALY approach assumes that a QALY has equal social value to everybody irrespective of the patient's age. However, it is possible that the general public assigns different social values to a QALY according to who gets it, derived from intergenerational equity judgements.

The aim of this paper is to discuss the possibility of weighting health gains by age in the aggregate QALY model. In the next section, we introduce ethical arguments in favor of using age–based weights. In the third section, the conditions under which it is possible to obtain a Social Welfare Function (SWF) —under certainty— consistent with the aggregate QALY model, are identified. One of these conditions is relaxed so that age–based weights can be introduced. In section four, we describe the design and the results of an experiment whose objective is to obtain a concrete functional form for SWF. This concrete form allows us to derive the age weights. Finally, section five discusses results obtained.

2 Some comments about age based equity weights

Aggregation in the QALY model, incorporates an equity judgement that some authors consider not only acceptable but desirable. Provided that one QALY has the same value whoever receives it, there is no discrimination amongst patients, beyond their capacity of improvement. But, is it a valid equity criterion?, Are there some other criteria that better approach social preferences? If this is the case, how can this criteria be introduced in QALY calculations?

Some authors have tried to answer these questions. Wagstaff (1991) reviews four definitions of equity, none of them are taken into account in QALY calculations, but they are frequently discussed in health economics literature. He selects the *equality of health* definition as the only one with a capacity of providing an adequate basis for determining an equitable allocation of resources. Starting from here, he proposes an isoelastic SWF that allows the introduction, not only of preferences for more equitable distributions, but to establish different weights for different patients. However, he does not provide any justification for the existence of age weights nor make a proposal for empirical estimation of this function and his weights.

Focusing on age weights, there are two main lines of thinking. One of them, is based on efficiency[1] (World Bank, 1993; Murray and Lopez, 1994; Murray, 1996). The other one is based on equity (Williams, 1997). Age weights based on efficiency emphasize the relationship between patient age and his social role. Children and, frequently, older patients, need physical and financial support from the others. Middle-aged adults are those who support this burden and therefore who contribute more to the social welfare. Based on that, Murray (1994) suggests *"unequal age–weights as an attempt to capture different social roles at different ages"*.

According to Williams, 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 methodology, younger patients will have higher weights, because they have completed a small portion of this normal span.

Each arguments has very different implications. Based on Murray's reason-

[1] Although, there are many meanings of efficiency, in this case we consider an allocation is more efficient when, at the same cost, it generates a higher social value. It must be distinguished from the notion of efficiency, which is implicit in the QALY model, where an allocation is more efficient when, at the same cost, a higher number of QALYs can be obtained.

ing, weights must be higher for middle–aged adults and decrease for younger and older people. However, Williams' reasoning implies monotonic decreasing weights with patient age.

On the other hand, some empirical studies have confirmed that some surveyed people take into account patient age when they are asked to evaluate social health programs. However, these studies are too restrictive. Most of them, do not obtain a quantifiable weight[2] (Nord, Richardson et al, 1995), and those that do, do not consider the possibility that weights depend on health gains that are being evaluated (Nord, Richardson et al, 1996; Busschbach and Hessing 1993).

The main contribution of our paper is that both the theoretical formulation and survey design are sufficiently flexible to allow the introduction of different age weights depending on health gains. *Ex ante*, there is no reason to think that these weights are independent of health gains, as Murray's formulation, as well as Williams's, implicitly suppose.

3 QALY model

3.1 Previous remarks

Before formalising the aggregate QALY model, it might be useful to make some clarifications. As we said in the introduction, a QALY is a health output measure. However, given that its theoretical foundations come from Utility Theory (Torrance, 1976; PlisKin, Shepard et al, 1980; Torrance and Feeny 1989; ...), most of the literature takes it as a utility measure. Although at an individual level, this difference is not relevant, at a social level it becomes important. If we consider a QALY as a health output measure, the aggregation of individual QALYs to get the aggregate measure of output produced by a health care program is trivial. However, if a QALY is considered a utility measure, we have the commonly known problem of interpersonal comparisons of utilities.

[2]The surveyed people answer questions as: "to implement a health program, do you think that younger people must be prioritised against older people, or that both groups must be treated with the same priority?

Starting from Sen (1970, 1977), Bleichrodt (1996) suggests that if we want to incorporate equity considerations in the cost–utility analysis based on QALYs, utility has to be considered as a cardinal measure and fully comparable. Therefore, the first question we must worry about is to what extent a QALY is a cardinal measure. Torrance (1986) identifies different methods to obtain QALYs that allow us to interpret them as cardinal measures —*standard game* in an uncertain world and *temporal exchange* in a certain world. To make interpersonal comparisons of utilities he proposes to assign zero value to the death state and one value to the full health state, and considers that the full health state has the same value for everybody.

Having defined a QALY as a cardinal measure and comparable between patients, we can identify the aggregation assumptions that are implicitly present in the aggregate QALY model in a certain world. The decision of treating the problem under certainty must be justified, because modelling under uncertainty is alternatively used by some authors.[3] First, we want to emphasize that, from a social point of view, uncertainty is a relative concept. Although at an individual level, results from health care programs are uncertain, at an aggregate level, results can be obtained with enough certainty if available information about success probability is not too imperfect.

Second, when we analyze the process of social decision making, the choice of a certain or an uncertain context is crucial because social preferences obtained by experimental methods under each environment can be very different. For example, risk aversion has a great influence on choices under uncertainty when these choices are modelled using the *"veil of ignorance"* approach.[4] However, when the choice under study is about different distributions of certain results, and this choice is made by an impartial observer which is not

[3] For instance, Bleichrodt (1996) identifies aggregation assumptions under uncertainty, using von Neumann-Morgenstern utility functions (von Neumann and Morgenstern, 1944) to represent preferences and considering, therefore, that preferences verify the axioms of Expected Utility Theory.

[4] Under this assumption, an agent must declare his preference amongst different alternatives, in our case health care programs, taking into account that he has the same probability of being any particular individual in society (Harsanyi,1955). The advantage of this assumption is that it allows us to obtain impartial preferences for different social alternatives.

affected by his decision, equity considerations are commonly incorporated into the analyses. This is the reason why, if the purpose is to study the social choice among outcomes, it is more appropriate to leave aside the influence of risk aversion, that affects the choices under uncertainty and, instead, introduce the equity judgements that are present in society.

3.2 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 patients and denote by $T \in \Re_+^n$ the set of possible results from different health care programs. An element of T is defined by an $n \times 1$ vector, $\tau = (t_1, \ldots, t_n)$, where $t_i \in \Re_+$, $i = 1, \ldots n$ refers to the number of life–years in full health[5] that are received from a health care program by patient i.

Suppose that we define a preference relation for patient i over T, by $\succeq_i$, with $\succeq_i$ meaning "*at least as preferred as.*" If $\tau \succeq_i \tau'$ and $\tau' \succeq_i \tau$, then τ is indifferent to τ', $\tau \sim_i \tau'$. If the relation $\succeq_i$ is complete, reflexive and transitive, (Debreu, 1954), it can be represented by means of a (individual) value function, $v_i(\tau)$, so that the expression $\tau \succeq_i \tau'$ is equivalent to $v_i(\tau) \geq v_i(\tau')$. Furthermore it is supposed that the relation is individualistic in the sense that preferences of patient i over the set T depend only on the individual outcome that he receives from each coordinate of τ, $v_i(\tau) = v_i(t_i)$, and that $v_i(t_i)$ is strictly increasing, that is, the patient always prefers more life–years in a healthy state.

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

$$U(\tau) = \sum_{i=1}^{n} u(t_i), \quad (1)$$

[5]In the QALY approach, the health gain can be measured as the number of quality–adjusted years. If the necessary assumptions to obtain QALYs are verified, working with QALYs is equivalent to working with life-years in full health. Hereafter, we will work with health gains measured in number of life-years in full health. This will simplify the experiment that we will perform in section 4.

where $u(t_i)$ is the social value function associated with individual result, $t_{i,}$ derived from a health care program.

Next we will derive under what conditions social preferences can be represented by this value function, in an certain world. To do this we consider a social planner (SP) and we consider that this SP is benevolent —he looks for social welfare— and impartial —he is indifferent to the identity of patients affected by his decisions. We consider that the SP's preferences will be a good approximation of society's preferences and, therefore, starting from the SP's preferences, we will obtain a Social Welfare Function (SWF).

To declare his preferences about different health care programs, the SP takes into account the individual welfare provided by each program. We want to get the preference relationship, $\succeq$, of the SP over elements in the set T, starting from the vector whose elements are the individual value functions defined on T. That is, we need to obtain a preference relationship over $v(\tau)$, where $v(\tau) = (v_1(t_1), \ldots, v_n(t_n))$. The preference relationship of the SP over the set $v(\tau)$ is complete, reflexive and transitive and then, it can be represented by a (social) value function —SWF— that we denote by $V(v(\tau))$. Our assumptions imply that $v(\tau) \succeq v(\tau')$ is equivalent to $V(v(\tau)) \geq V(v(\tau'))$. In addition, SP's preferences depend positively on individual preferences. That is, if $\tau \succeq_i \tau'$ for some patient i, and $\tau \sim_i \tau'$ for the rest of patients, then $\tau \succeq \tau'$ for the SP. Previous considerations are formalised as follows:

Assumption 1 The SP's value function is strictly increasing in $v_i(t_i)$ and therefore in t_i.

To obtain the additive function of the aggregate QALY model, we need to make an assumption of independence. Suppose that we divide the coordinates of vector $v(\tau) = (v_1(t_1), \ldots, v_n(t_n))$ into two subvectors $(v_{ij,}\overline{v}_{ij})$, where $v_{ij} = (v_i(t_i), v_j(t_j))$ and $\overline{v}_{ij}$ is its complement.

Definition 1 The vector v_{ij} is preferentially independent of its complement $\overline{v}_{ij}$, if and only if the preferences over v_{ij}, given $\overline{v}'_{ij}$, don't depend on $\overline{v}'_{ij}$, for every $\overline{v}'_{ij}$. That is,

$$\left(v'_{ij}, \overline{v}'_{ij}\right) \succeq \left(v''_{ij}, \overline{v}'_{ij}\right) \Rightarrow \left(v'_{ij}, \overline{v}_{ij}\right) \succeq \left(v''_{ij}, \overline{v}_{ij}\right), \qquad all \overline{v}_{ij}, v'_{ij}, v''_{ij}.$$

It seems suitable to consider that the SP's preferences over two different distributions of health, that are different only in the health provision of two patients, depend only on preferences of the SP over the preferences of those two patients. Formally stated,

Assumption 2 The vector v_{ij} is preferentially independent of its complement $\overline{v}_{ij}$ for all $i \neq j$ and $n \geq 3$.

These assumptions allow us to postulate an additive value function for the SP using the following theorem:

Theorem 1 (Debreu, 1960; Gorman 1968). SP value function, V, is additive,

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

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

Where v_i^* is a positive monotonic transformation defined over v_i, and reflects the interpersonal comparisons of utilities made by the SP.

It must be taken into account that in the QALY model defined in equation (1), $v_i^*\left(v_i\left(t_i\right)\right)$ is common to all patients. Therefore, it is necessary to make an additional assumption to obtain a complete characterisation of QALYs. To do that, we introduce the concept of permutation.

Let $\tau = (t_1, \ldots, t_n)$, be a distribution of health. We define a permutation of τ, and denote it by σ, as an exchange of individual values in the set of patients, $\{1, \ldots, n\}$. Then, $\tau_\sigma = \left(t_{\sigma(1)}, \ldots, t_{\sigma(n)}\right)$ is a permutation, where $t_{\sigma(i)}$ is the permuted value of patient i, $i = 1, \ldots, n$.

Assumption 3. Anonymity. $\tau \sim \tau_\sigma$ for all health distribution τ and for all permutation σ.

This assumption says that, if a distribution of health gains is a permutation of a given health gains distribution, 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 life–years in full health received by i and the second to those received by j.

Condition 1 Given that $v_i^*(v_i(t_i))$ is increasing in t_i and $t_i \geq 0$, then $t_i = 0$ is the number of health gains least preferred by both, individually and by the SP. Therefore, we will consider that the social value of $t_i = 0$ is the same and equal to zero, $v_i^*(v_i(0)) = 0$ for all i.

Proposition 1 Under assumptions 1, 2 and 3, and condition 1, we have that, $v_i^*(v_i(t_i)) = v_j^*(v_j(t_j))$, for all i, j and $t_i = t_j$.

A proof can be found the appendix A.

This proposition guarantees that a given health gain will have the same social value independently of who receives it. Moreover, it allows us to characterise the social value function used in the aggregate QALY model under certainty, and previously defined in equation (1).

$$V(v(\tau)) \equiv U(\tau) = \sum_{i=1}^{n} u(t_i) \quad where u(t_i) = v_i^*(v_i(t_i)) \quad all i$$

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

3.3 Weighted QALY Model

Suppose that the SP's preferences —and, therefore, the SWF—, depend, not only on individual preferences, but on patient age, also. Under this assumption, we want to get the SP preference relation, $\succ$, defined on the set $(E, V(T))$. One element on this set $(E, V(T))$, that we denote as $(\epsilon, v(\tau))$, can be obtained starting from an element $v(\tau) = (v_1(t_1), \ldots, v_n(t_n))$ through $((e_1, v_1(t_1)), \ldots, (e_n, v_n(t_n)))$, where $\epsilon = (e_1, \ldots, e_n) \in \Re_+^n$ is the ages vector.

Given that the SP preference relation over $(E, V(T))$ is complete, reflexive and transitive, it can be represented by means of a value function that we denote by $W(\epsilon, v(\tau))$ such that, $(\epsilon, v(\tau)) \succeq (\epsilon', v(\tau'))$ is equivalent to $W(\epsilon, v(\tau)) \geq W(\epsilon', v(\tau'))$.

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 $v_i(t_i)$, given ϵ. Assumption 2′

states that the SP's preferences between two health distributions that are different only in health gains received by two patients, depend only on individual preferences of both patients and on interpersonal comparisons of utility, based on ages, made by the SP. That is, assumption 2 is maintained but now taking into account that $v_{ij} = ((e_i, v_i(t_i)), (e_j, v_j(t_j)))$.

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

$$W(\epsilon, v(\tau)) = \sum_{i=1}^{n} w_i(e_i, v_i(t_i)),$$

where w_i is a positive monotonic transformation of v_i and reflects the SP's interpersonal comparisons of utility but taking into account patient age.

Now, we adopt the additional assumption that a given increase in health level, has the same value for patients of the same age. Formally, given any distribution of health gain, $\tau = (t_1, \ldots, t_n)$, let δ denote a permutation of individual values of any patients of the same age in the set $\{1, \ldots, n\}$, then,

Assumption 4. *Conditioned anonymity.* $\tau \sim \tau_\delta$ for any distribution τ and any permutation δ.

This weakens the anonymity assumption that is implicitly present in the aggregate QALY model, defined in equation (1). Now, an increment can have a different value depending on patient ages.

Condition 1′ Given an age, e_0, that we call "reference" age, $t = 0$ is the health gain least preferred by the SP, as well as by each patient. We consider that the SP's valuation of zero gain by patients of reference age, is zero. That is, $w_i(e_0, v_i(0)) = 0$ for all i such that $e_i = e_0$.

Proposition 2 Under assumption 1′, assumption 4 and condition 1′, $w_i(e_i, v_i(t_i)) = w_j(e_j, v_j(t_j)) \equiv w(t^*)$, for all $e_i = e_j \equiv e_0$ 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, e_0.

Given that the value function is the same for patients of the same age, it is possible to reduce the dimensionality of our problem, transforming increments of health received by patients of different ages into increments of health in reference age.

Let us fix as a reference age the age of patient h, $e_0 = e_h$, and suppose that, for either $(e_i, v_i(t_i))$, there exists a value, $t^*_{(i)}$, that makes the SP indifferent to allocations $(e_i, v_i(t_i))$ and $\left(e_0, v_h(t^*_{(i)})\right)$. Then $w_i(e_i, v_i(t_i)) = w_h\left(e_0, v_h(t^*_{(i)})\right) \equiv w\left(t^*_{(i)}\right)$.

Therefore, starting from the aggregate QALY model, we have obtained the following model, which we call weighted QALY model,

$$W(\epsilon, v(\tau)) = \sum_{i=1}^{n} w_i(e_i, v_i(t_i)) = \sum_{i=1}^{n} w(t^*_{(i)}), \qquad (2)$$

where $t^*_{(i)} = g(e_i, t_i; e_0)$.

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 in the sense that the latter can be obtained as a particular case of the former when age weights are equal to everybody. That is to say, $g(e_i, t_i; e_0) = t_i$. Note that we do not impose any functional form to the function $g(\cdot)$. This allows the weights assigned to each age to vary, as a function of health gains. Although this is a nice property of our model, it brings a difficulty when we try to implement it empirically: we have infinite combinations of age and health increments. One possible solution is to select a finite —and small— number of combinations, that we consider more representative *ex ante*; to obtain the value of t^* for each one; to look for a functional form that fits the observed values reasonably well and, finally, to use this functional form to extrapolate the values of any other combination. This is the procedure we use in the experiment we are going to show in the next section.

4 The experiment

4.1 Design

To analyze if patient age is a relevant variable in the social valuation of health care programs, we did an experiment with 61 undergraduate students —21 of Economics, 20 of Political Science and 20 of Law. 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. Then, they answered a pilot questionnaire to familiarise themselves with the type of questions that they would have to answer in the following 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 (e, t), where e is the patient age, and t is the health gain, measured in healthy live–years, provided to this patient. For each program, each participant had to decide the increment of years, t^*, that would make him indifferent between the proposed program (e, t) and the program $(20, t^*)$, where $e = 20$ is the reference age. This is known as the matching technique.

In the pilot questionnaire we detected that the participants had some difficulties to choose on an specific number t^*, so we decided to use the "balance mechanism". It consists of approaching the value through successive questions where it is always necessary to choose between two allocations —see appendix B.[6]

The construction of a list of health care programs for evaluation was not easy. Some empirical studies have detected a trade–off between the number of programs to be evaluated and 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

[6] Sometimes the balance mechanism doesn't allow to get an exact value for t^* but an interval. In this case we take the intermediate value of that interval.

$(e, t) = (60, 40)$ 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 age and the second refers to health gain given to this patient.

After applying the balance mechanism to all programs on the list, we provided each participant with fifteen cards that they had to rank from more to less preferred. Each card contains each one of the programs evaluated previously. 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. The purpose now was to repeat some of the tasks of the previous meeting. We divided the participants into three groups of 20 people in each one —one of the previous participants did not come. Each group matched four of the eleven programs —a program was evaluated by two groups— and ranked five of fifteen possible cards.

4.2 Method of analysis

First, from 61 valuations of participants obtained with the matching technique, we calculate the average value for each one of eleven combinations of age and health gains. We verify if average valuations are significantly different across different combinations. We specially look at health care programs that give the same gain to patients of different ages, because these gains should have the same value if participants do not take into account age and, therefore, if the assumptions of the QALY model, defined in equation (1), were verified.

Once we have obtained the average valuation for each one of the chosen combinations, we want to get a function that allows us to recover the values of those combinations that are not evaluated directly. Given that we do not want to impose any assumption on the function, we estimate different functional forms using Ordinary Least Squares until we find that one that best fits the data.

Some authors (Streiner and Norman, 1991), argue that in an experiment of this nature it is important to check if our technique is measuring what we want —validity—, and if the results are consistent across time— temporal reliability. To verify the validity of our method, we have some difficulties. Obviously, we do not know what the preferences are and, therefore, we cannot verify if we are measuring them correctly. However, we can approximate the validity test checking if a different method gives similar results. For this reason we utilise, in addition to the matching technique, the direct ordering technique. It is true that the ordering technique does not allow us to obtain cardinal values, but it probably reflects a fairly accurate approximation of preferences. This possibility makes it interesting to compare these orderings to those obtained with the matching technique at an individual level —correlation among the answers of a same participant— as well as at a social level —correlation among the two aggregate orderings assessed. Anyway, comparing both techniques we will be able to analyze the consistency across methods.

Spearman rank correlation coefficient (SCC) is applied to evaluate the correlation between both orderings at an individual level. Then, we calculate the average SCC of all participants. To analyze the correlation at a social level, we apply SCC, as well as Kendall rank correlation coefficient (KCC).[7] But, previous to these calculations, we need to aggregate individual preferences obtained using both techniques. Given that direct ordering only gives information about order, we use the Borda rule to this end. However, to calculate the social ordering starting from individual matching, we use two methods. The first, which we call *Borda–matching,* transforms the individual answers —that are cardinal— into ordinal measures (ranking) that are then aggregated using the Borda rule. The second, which we call *direct–matching*, directly adds up the t^* values given by participants to each combination. It must be taken into account that the degree of coincidence of social rating from direct ordering and each one of social ratings from matching gives us different information. If we adopt *Borda–matching*, we have a measure of the consistency across methods, and/or the lack of precision in the people

[7]SCC and KCC are non–parametric techniques which are applicable to ordinal date. Their values lie between -1 and 1, higher value indicating stronger positive association between the ranks.

preferences, because both answers should be the same.[8] Moreover, as we previously mentioned, it can be considered as a good approximation to test validity. However, if we aggregate using *direct–matching*, we are measuring, in addition, preference intensity. Even in the hypothetical case that all surveyed people give the same ordering using matching and direct ordering, only with the first technique can we guarantee identical social orderings.

Finally, we study temporal reliability (retest) for each one of the participants. To do that, we analyze the degree of coincidence between initial answers and those obtained two weeks after, with each one of the two techniques. To calculate the correlation between initial and final direct ordering, we compute SCC between the ranking of five combinations from the retest and the ranking that the same cards had in the initial survey. To calculate the correlation between initial and final values, we obtain Pearson linear correlation coefficient (PCC) among five values from the retest and the values that the same combinations had in the initial survey.[9]

4.3 Results

Table 1 shows average values, from 61 participants, for each combination of patient age and healthy life–years gain obtained using the matching technique. It should be noted that the right column indicates the additional life-years number that an individual 20 years old should receive such that the mean participant is indifferent to this treatment and the left column treatment. Once we have the average valuation, and before making some comments about results, we must carry out some contrasts to know if these valuations are significantly different from one another. In order to do this, a mean difference contrast is made, excepting when we compare the value of any gain for ages other than twenty, with the value of any gain assigned to

[8]The lack of precision of preferences is extremely relevant when individual preferences are analysed (Dolan, 1997; Fischhoff, 1991). People cannot be expected to have articulated opinions on more than a small set of issues —of which the subject of this experiment is unlike to be one— with which they are very familiar. Therefore, some of the inconsistencies that individuals produce can appear because they are formatting —and probably changing— their preferences as they are filling out the questionnaire.

[9]PCC is a parametric technique which is applicable to cardinal date. His value lie between -1 and 1, higher value indicating stronger positive association between the ranks.

20–year–old patients. In this case we use a mean contrast.

Table 1: Social value of health gains based on age[1]

Age (e), Health gain(t)	$(t--Student) Social value$(t*)
20,40	40
1,40	33.4
	(-3.19)
40,40	29.1
	(-6.55)
20,20	20
40,20	18.6
	(-1.03)
1,20	13.2
	(-5.55)
60,20	10.9
	(-10.58)
40,10	10.1
	(0.20)
20,10	10
60,10	6.3
	(-8.23)
1,10	5.4
	(-7.45)
20,2	2
40,2	1.9
	(-0.41)
60,2	1.4
	(-2.93)
1,2	0.9
	(-8.41)
(1)All variables are measured in years	

Table 1 only shows t–Student statistic that allows us to contrast if the average value of t^* is significantly different from t (mean contrast), that is to say, it only analyses treatments which are directly compared. As can be seen, all values of t^*, but three, are different at 95% confidence level. Therefore, although the average value is different, the hypothesis that gains of 2, 10 and

20 years are equal when 20–year–old and 40–year–old patients are compared is not rejected. Although we do not show it, the rest of contrasts have displayed that the only pair of combinations whose averages are different at 95% confidence level, are $(60, 20)$ and $(40, 10)$.

With regard to results, table 1 shows that the valuation of the four gains considered is strongly conditioned by patient's age. Moreover, the age influence varies with the gain we are considering. When the gain is of 40 life–years, 20–year–old patients and 1–year–old patients are ranked above 40–year–old patients. However, for smaller gains, 20–year–old patients and 40–year–old patients are ranked first. It must be emphasised that children go down in ranking as the number of life–years provided decreases. This can be more clearly observed if we calculate age–based relative weights, assigned to each gain, starting from average valuations of table 1. To do this, we assign an arbitrary weight and equal one to any gains received by 20–year–old patients —reference age.

Figure 1 shows weights received by different gains are similar when middle–aged and young patients are considered, except if the gain is of 40 years. On the contrary, the weight assigned to children starts being very small, and also considerably smaller than those assigned to 60–year–old patients, but it increases very fast as the gain is increased. Weights assigned to 60–year–old patients decrease very quickly as the gain is increased.

We have obtained the weights that surveyed people give to four different gains taking into account four possible patient ages. But, what can we say for other possible combinations? How can we know, for instance, if a gain of 15 years of life for a 50–year–old patient is more valued than a gain of 8 yeas of life for a 30–year–old patient? To answer these questions, we estimated different functional forms using Ordinary Least Squares, starting from individual results obtained with matching. Although we are aware that the information obtained in the survey is limited, the purpose is to look for a function that fits well with the preferences of the people surveyed in the sample. In terms of the weighted QALY model, we want to approximate the function $g(\cdot)$.

We choose a quadratic function without constant because it fits better our data. The estimated equation is the following,

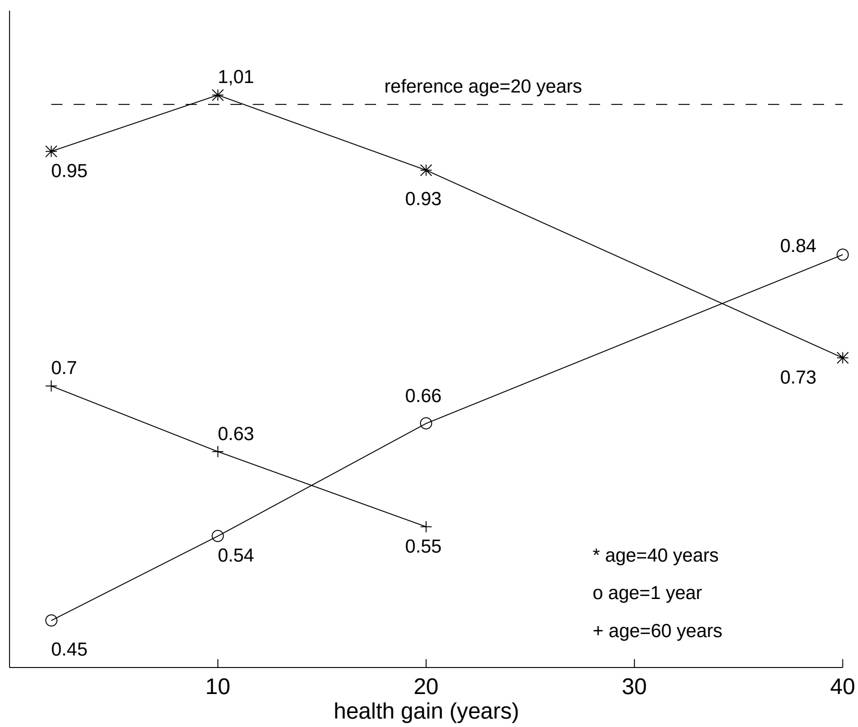

Figure 1: Weights of health gains based on age

$$
\begin{aligned}
\hat{g}(e,t;e_0) &= 0.263 * e + 0.684 * t - 0.004 * e^2 + 0.004 * t^2 - 0.003 * e * t \\
&\quad\ (7.42) \qquad\quad (12.9) \qquad\quad (-7.46) \qquad\quad (3.22) \qquad\quad (-3.6) \\
R^2 &= 0.72
\end{aligned}
$$

where e is the patient age, t is the health gain, and $\hat{g}(e,t;e_0)$ is the estimated valuation for each combination (e,t).
We use this function to obtain the valuation of all combinations not directly evaluated in the survey. For example, we can say that a health gain of 15 years in a 50–year–old patient is preferred to a gain of 8 years in a 30–year–old patient.

Starting from partial derivatives of $\hat{g}(\cdot)$,

$$
\begin{aligned}
\hat{g}'_e(\cdot) &= \delta\hat{g}(\cdot)/\delta e = 0.263 - 0.008 * e - 0.003 * t \\
\hat{g}'_t(\cdot) &= \delta\hat{g}(\cdot)/\delta t = 0.684 + 0.008 * t - 0.003 * e,
\end{aligned}
$$

it is possible to analyse to what extent the health program valuation depends on both variables, the patient age and the life–years gain. On one hand, given t, the fact that $\hat{g}'_e(\cdot)$ can be different from zero, shows that changes in patient age can affect the program assessment. Therefore the age is a relevant variable. Besides, this variation is not constant, being even possible to have different sign, depending on values of t and e. On the other hand, given e, $\hat{g}'_t(\cdot)$ is positive for all feasible ages and life–years gain. As it was expected a life–years gain always increases the utility, no matter what the age is.

Now we study the validity of the results. As we have seen, the analysis of consistency across methods can be a good approximation to a validity study. Table 2 presents the social ordering, from more to less preferred, obtained using direct ordering (DO), direct–matching (DM) and Borda–matching (BM).

Table 2: Ranking of health care programs[1]
(from more to less preffered)

Age (e), Gain(t)		
DM	BM	DO
20,40	20,40	20,40
1,40	1,40	40,40
40,40	40,40	1,40
20,20	20,20	20,20
40,20	40,20	40,20
1,20	1,20	20,10
60,20	20,10	1,20
40,10	60,20	40,10
20,10	40,10	60,20
60,10	60,10	60,10
1,10	1,10	1,10
20,2	20,2	20,2
40,2	40,2	40,2
60,2	60,2	60,2
1,2	1,2	1,2

(1)All variables are measured in years
CCK=0.92(MD vs OD); 0.94(MB vs OD)
CCS=0.94(MD vs OD); 0.99(MB vs OD)
CCS individual (average)=0.79

As we have mentioned to analyze consistency across methods it is necessary to compare results from DO with results from BM. It can be seen that results from both methods present a higher correlation. Therefore, we can say that both techniques provide similar orderings and this gives a high level of confidence in the results. Furthermore, on the other hand, DO and DM provide a high correlation also. This means that, even if we incorporate preference intensity into the analysis, we get very similar social orderings.

Finally, we analyze the reliability of the results studying temporal consistency. For each participant we calculate SCC between initial and final direct ordering (retest), and PCC between initial and final values obtained with matching technique. We obtain an average SCC value of 0.89 and an average

PCC value of 0.93. This shows a high stability of preferences through time and supports the reliability of the results.

5 Discuss and conclusions

In this paper, we have shown that the aggregate QALY model under certainty, commonly used in empirical studies, can be derived from some customary assumptions in social choice literature. It was also shown that replacing the assumption of anonymity by conditional anonymity, enables the possibility that health gains can be weighted by age. One of the advantages of the proposed methodology is its flexibility and its easy empirical implementation. By comparing health gain at any age with health gain at a reference age, we can obtain relative based–age weights, that can be used in economic evaluation of health care.

The results of the experiment reported in this paper show that the respondents consider that the patients age is a relevant factor when they have to assess health gains. Moreover age weights are not constant, but they vary depending on gains considered. For example, whereas the relative weight of infants rises with health gain, the weight of old people decreases. This result leads us to reject the QALY model, defined in equation (1), as a good representation of the preferences sampled. On the contrary, the weighted QALY model —equation (2)— gives us a better fitting of this preferences.
From relative weights obtained it is interesting to examine which age weight argument —efficiency or equity— provides a better description of respondents' preferences. In principle, middle–aged patients have a high weight, independent of the health gains. This supports the social efficiency argument. However, the span that the weight of infants follows, provides additional information. So, as the gain is increased, his weights increase —it is even possible that infants would have the highest weight, for gains greater than 40 life–years. It seems to reveal a reinterpretation of the equity criteria: The lower the age of a person relative to the average life expectancy, the higher its weight, provided that with the treatment the person reaches a "reasonable" age.

The interviews carried out have the advantage that they allow us to get at the same time qualitative and quantitative information. So, asking participants to write a brief explanation of their answers helped to understand results. When gains considered are small, most of the participants assign high weights to middle–aged patients and small weights to children, mainly due to two reasons: first, because it is likely that middle aged patients have family responsibilities and, second, because a small health gain to a infant is not enough to achieve some targets contained in what we call "establishing one's own personality" —psychological maturity, becoming an independent person, establishing a home, profiting from acquired knowledge ... However, as health gain is increased, the justification for answers are diverse. On one hand, some people justify their valuations based on equity reasons and, therefore, they assign priority to children and young patients. On the other hand, some other surveyed people give more importance to social efficiency so they give less weight to children. Since both criteria go in the opposite direction, the average weight given to children is high when a big gain is considered, but is not the highest. However, as we mentioned, as the gain considered increases, participants give a higher relative weight to children and a smaller one to middle–aged adults. That is, the equity argument becomes more important than the social efficiency argument.

Finally, we would like to project the analysis of validity and reliability. We have seen that using an alternative technique similar results are obtained. This supports the validity of the proposed technique. Moreover, reliability is also supported because when we partially repeated the experiment two weeks later, the results are scarcely modified.

To evaluate the results, limitations of our sample must be taken into account —age and number of participants among others things. However, in spite of this, we have covered the initial purpose of this paper and established a good starting point for future work.

Appendix A

Let $v_i^*[v_i(t)]$ be denoted as $u_i(t)$ for all i. 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, we have,

$$u_h(t_h) + u_p(t_p) + \overline{u}_{hp} = u_h(t_p) + u_p(t_h) + \overline{u}_{hp} \quad foranyt_h,t_p$$

where, $\overline{u}_{hp} = \sum_{i=1}^{n} u_i(t_i) - u_h(t_h) - u_p(t_p)$.

Given that $[u_h - u_p](t_h) = [u_h - u_p](t_p)$, then, $[u_h - u_p](t) = k$ for all t. Where k is a constant.

From condition 1, $[u_h - u_p](0) = 0$. Starting from the last two equalities we obtain $u_h(t) = u_p(t)$, 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 increase of healthy life–years that are provided to the patient, and in the patient age who receive gains. You must say whether you prefer treatment A, treatment B, or you are indifferent to both. Depending on your choice the questionnaire continues in the following way:
– If you choose an option where you find 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 where you find 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:

Treatment A		Treatment B				
Age	*Years*	*Age*	*Years*	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

References

[1] Bleichrodt, H. (1996). *Applications of utility theory in the economic evaluation of health care.* Thesis Erasmus University Rotterdam.

[2] Busschbach, J.J.V.; Hessing, D. J.; Charro, F. T. (1993). "The utility of health at different stages in life: A quantitative approach." *Social Science and Medicine*, 37(2):153-158.

[3] Debreu, G. (1954). *Representation of a preference ordering by a numerical function.* R. M. Thrall, C. H. Coombs and R.L. Davis, eds. Decision Processes. Wiley, New York, 159-165.

[4] Debreu, G. (1960). "Topological methods in cardinal utility theory." K. J. Arrow, S. Karlin and P. Suppes, eds. *Mathematical methods in the social sciences.* Stanford University Press, 16-26.

[5] Dolan, P. (1997). "The nature of individual preferences: a prologue to Johannesson, Jonsson and Karlsson". *Journal of Health Economics*, 6: 91-93.

[6] Fischhoff, B. (1991). "Value elicitation: Is there anything in there?". *American Psychologist*, 46: 835-847.

[7] Gorman, W. M. (1968). "The structure of utility functions." *Review of Economic Studies*, 35: 367-390.

[8] Harsanyi, J.C. (1955). "Cardinal Welfare, individualistic ethics, and interpersonal comparisons of utility." *Journal of Political Economy*, 63, 309-321.

[9] Loomes, G.; McKenzie, L. (1989). "The use of QALYs in health care decision making." *Social Science Medicine*, 28(4): 299-308.

[10] Murray, C. J. L.; Lopez, A. D. eds.(1994). *Global Comparative Assessments in the Health Sector: Disease Burden, Expenditures and Intervention Packages*, WHO.

[11] Murray, C. J. L.; Acharya, A.K (1997). "Understanding DALYs". *Journal of Health Economics, 16:703-730.*

[12] Nord, E.; Richardson, J.; Street, A.; Kuhse, H.; Singer, P. (1995). "Maximizing health benefits versus egalitarianism: An Australian survey of health issues." *Social Science and Medicine*, 41(10): 1429-1437.

[13] Nord, E.; Richardson, J.; Street, A.; Kuhse, H.; Singer, P. (1996). "The significance of age and duration of effect in social evaluation of health care." *Health Care Analysis*, 4: 103-111.

[14] Olsen, J.A. (1994). "Persons vs years: two Ways of eliciting implicit Weights." *Health Economics*, 3: 39-46.

[15] PlisKin, J. S.; Shepard, D. S.; Weintein, M. C. (1980). "Utility functions for life years and health status." *Operations research*, 28: 206-24.

[16] Sen, A. K.(1970). *Collective choice and social welfare.* Holden-Day, San Francisco.

[17] Sen, A. K.(1977). "On weights and measures: Informational constraints in social welfare analysis." *Econometrica*, 45: 1539-1572.

[18] Streiner, D; Norman, G.R. (1991). *Health measurement scales. A practical guide to their Development and use.* Oxford University Press.

[19] Torrance, G. W. (1976). "Toward a utility theory foundation for health status index models." *Health Services Research*, winter, 349-369.

[20] Torrance, G. W. (1986). "Measurement of health state utilities for economic appraisal: a review." *Journal of Health Economic*, 5:1-30.

[21] Torrance, G. W.; Feeny D. (1989). "Utilities and quality-adjusted life years." *International Journal of Technology Assessment in Health Care*, 5:559-575.

[22] Tsuchiya, A. (1997). "Age-related Preferences: Should Health Benefits be Age Weighted?" (submitted for publication).

[23] Ubel, P.; Loewenstein, G.; Scanlon, D.; Kamlet, M.(1996). "Individual utilities are inconsistent with rationing choices: a partial explication of why Oregon's cost-effectiveness list failed." *Medical Decision Making*, 16:108-116.

[24] von Neumann J.; O. Morgenstern (1944). *Theory of Games and Economic Behavior*. Princeton University Press, Princeton.

[25] Wagstaff, A. (1991). "QALYs and the equity-effiency trade-off." *Journal of Health Economics* 10, 21-41.

[26] Williams, A.(1997). "Intergenerational equity: an exploration of the 'fair innings' argument." *Health Economics*, 6:117-132.

[27] Word Bank (1993). *World Development Report 1993: Investing in Health*. Oxford University Press.