Available online at www.sciencedirect.com
SCIENCE DIRECT
ELSEVIER
Journal of Health Economics 24 (2005) 655–678
JOURNAL OF HEALTH ECONOMICS
www.elsevier.com/locate/econbase

A nonparametric elicitation of the equity-efficiency trade-off in cost-utility analysis

Han Bleichrodt[a,*], Jason Doctor[b], Elly Stolk[a,1]

[a] iMTA/iBMG, Erasmus University, P.O. Box 1738, 3000 DR Rotterdam, The Netherlands
[b] University of Washington, Seattle, WA, USA

Received 11 September 2003; received in revised form 26 August 2004; accepted 18 October 2004
Available online 25 December 2004

Abstract

We performed an empirical elicitation of the equity-efficiency trade-off in cost-utility analysis using the rank-dependent quality-adjusted life-year (QALY) model, a model that includes as special cases many of the social welfare functions that have been proposed in the literature. Our elicitation method corrects for utility curvature and, therefore, our estimated equity weights are not affected by diminishing marginal utility. We observed a preference for equality in the allocation of health. The data suggest that the elicited equity weights were jointly determined by preferences for equality and by insensitivity to group size. A procedure is proposed to correct the equity weights for insensitivity to group size. Finally, we give an illustration how our method can be implemented in health policy.

JEL classification: D63; I10

Keywords: Equity-efficiency trade-off; Cost-utility analysis

1. Introduction

The common procedure to aggregate health benefits in economic evaluations of health care is by unweighted aggregation, also referred to as quality-adjusted life-year (QALY)-

* Corresponding author. Fax: +31 10 408 9092.
E-mail address: bleichrodt@bmg.eur.nl (H. Bleichrodt).
[1] The order of authorship is alphabetical to reflect equal contributions.

doi:10.1016/j.jhealeco.2004.10.001

utilitarianism. This procedure weights the health gains of each individual equally and leads to a maximization of health gains. Several authors have raised concerns about the equity implications of QALY-utilitarianism and have argued that it may be necessary to differentiate between individuals based on, for example, age, health status or previously enjoyed health (Harris, 1987; Nord, 1995; Williams, 1997; Williams and Cookson, 2000).

Empirical evidence supports these concerns and indicates that people, when choosing between different allocations of health gains, not only consider efficiency, the total amount of health gains, but also equity, the distribution of the health gains (e.g. Nord, 1993; Dolan, 1998; Abellan and Pinto, 1999). These findings suggest that it may be preferable to replace QALY-utilitarianism by some sort of equity-weighted aggregation rule. Unfortunately, the available empirical research offers little guidance as to which rule should be used and how the equity weights could be elicited.

Several authors have proposed theoretical models to incorporate equity considerations into cost-utility analysis (Wagstaff, 1991; Bleichrodt, 1997; Williams, 1997; Dolan, 1998). Both Wagstaff (1991, 1993) and Dolan (1998) proposed to use an iso-elastic social welfare function to allow for a trade-off between efficiency and equity. Within this class of social welfare functions, Dolan (1998) suggested, in particular, to use a Cobb–Douglas function. Wagstaff (1991) and Dolan (1998) did not derive the assumptions underlying their proposed social welfare functions, which complicates an assessment as to why the equity-efficiency trade-off should take the form they proposed. They did not explain either how the parameters in their social welfare functions could be assessed.

Bleichrodt (1997) proposed a multiplicative social welfare function, derived the conditions on which it depends, and showed how its equity parameter could be elicited. The range of equity concerns that the multiplicative social welfare function can address is, however, limited. Williams (1997) suggested that individuals should be weighted according to their ‘fair innings’, the difference between the amount of health they already enjoyed and the amount of health they are entitled to over their lifetime. Williams’ proposal suggests that he had in mind some sort of weighted aggregation rule, but he did not specify what form this weighted rule should take nor did he explain how the equity weights could be elicited.

Bleichrodt et al. (2004) recently proposed a new social welfare function to incorporate equity considerations into cost-utility analysis, the rank-dependent QALY model. Their model has several desirable characteristics. First, it is consistent with several social welfare functions that have been proposed in the literature, including QALY-utilitarianism, the Rawlsian social welfare function in which all weight goes to the worst-off individual, and the Gini social welfare function, which is widely used in inequality measurement. The rank-dependent QALY model can also accommodate Williams’ fair innings approach. Second, as Bleichrodt et al. (2004) showed, the rank-dependent QALY model depends on assumptions that have normative appeal. A third advantage of the model is that the elicitation of the equity weights is straightforward. Finally, the model is tractable: once the equity weights have been elicited, the model can easily be used in cost-utility analyses.

The aim of this paper is to elicit the equity weights under the rank-dependent QALY model. For reasons explained in Section 2, we used a more general model than the model proposed in Bleichrodt et al. (2004). In Bleichrodt et al. (2004), the social utility function over QALYs is linear, whereas in this paper, we allow for a nonlinear social utility function over QALYs. We refer to this extended model as the nonlinear rank-dependent QALY model. A consequence of using a more general model is that its elicitation becomes more involved, because, in addition to the equity weights, the social utility function over QALYs must be determined.

The structure of the rest of the paper is as follows. In Section 2, we describe the nonlinear rank-dependent QALY model. In Section 3, we explain the elicitation of the model. To elicit the model, we used an adjusted version of the trade-off method (Wakker and Deneffe, 1996), which was developed to measure utilities under risk. An advantage of the trade-off method is that it is nonparametric: it imposes no assumptions on the utility function or on the equity weighting function. We elicited the nonlinear rank-dependent QALY model both in a sample of students and in a sample of the general population. Section 4 describes the designs of the two experiments, Section 5 the results. Section 6 shows how our method can be implemented in health policy. Section 7 offers concluding remarks.

2. The rank-dependent QALY model

We consider a health policy maker who has to choose between different QALY allocations. Consider a population of n individuals. Let $(q_1, \ldots, q_n)$ denote the *QALY-profile*, which gives q_i QALYs to individual i. We will interpret QALYs as measures of health in this paper. Unless otherwise stated, we assume that QALY-profiles are *rank-ordered* so that $q_1 \geq \cdots \geq q_n$. This is, obviously, no restriction because each QALY-profile can be written in a rank-ordered form.

In this paper, we study preferences over QALY-profiles. To describe these preferences, Bleichrodt et al. (2004) suggested using the rank-dependent QALY model. According to the *rank-dependent QALY model*, the social value of QALY-profile $(q_1, \ldots, q_n)$ is equal to:

$$\sum_{i=1}^{n} \pi_i q_i, \tag{1}$$

where the π_i are equity weights that are defined as $\pi_i = w(i/n) - w((i-1)/n)$. The function w is a nondecreasing function that has $w(0) = 0$ and $w(1) = 1$.

Under the rank-dependent QALY model, the social value of a QALY allocation is thus expressed in terms of two scales, w and q. The scale q is the familiar one for quality-adjusted life expectancy. The other scale, w, associates with each individual's expected quality-adjusted lifetime, q_i, an equity weight π_i, which reflects the weight the policy maker gives to individual i in the evaluation of QALY-profiles.

Under the rank-dependent QALY model, the equity weight assigned to an individual depends on how well-off he is in terms of QALYs by comparison with the other individuals in society, i.e. the equity weight depends on the individual's rank. A shift in the individual's rank will generally lead to a shift in his equity weight. A detailed explanation of the intuition behind the rank-dependent QALY model is given in Bleichrodt et al. (2004).

It is easily verified that in case the function w is linear, the rank-dependent QALY model is identical to QALY-utilitarianism. If w is convex then the policy maker is averse to inequalities, in the sense that he will always prefer a transfer of QALYs from an individual who has relatively many QALYs to an individual who has less, as long as the rank-ordering of the individuals in terms of the number of QALYs received is not affected. If w is concave then the policy maker is inequality seeking. Because the function w describes attitudes towards inequality, we refer to this function as the *equity weighting function*.

In this paper, we consider a generalized version of the rank-dependent QALY model, the *nonlinear rank-dependent QALY model,* in which the value of the rank-ordered QALY-profile $(q_1, \ldots, q_n)$ is equal to:

$$\sum_{i=1}^{n} \pi_i U(q_i). \tag{2}$$

The difference with Expression (1) is that in the nonlinear rank-dependent QALY model the utility function U over QALYs need not be linear. An important point to note is that the utility function U in Expression (2) is the policy maker's utility function over QALYs; it reflects the value the policy maker places on different numbers of QALYs experienced by the people in society. Assuming the existence of a social utility function is common in the literature on inequality measurement (e.g. Atkinson, 1970; Ebert, 1988). For health, the approach of defining a social utility function over QALYs has been used by Wagstaff (1991), Bleichrodt (1997) and Dolan (1998). Note that Expression (2) could be made consistent with individuals valuing their *own* QALYs in a nonlinear manner by substituting $u_i(q_i)$ for q_i, where u_i is an individual utility function over QALYs. We do not pursue such an extension in this paper.

The reason to allow for nonlinear utility over QALYs is that the elicitation of social preferences is a descriptive task and it is not a priori clear that a linear utility function over QALYs describes preferences over QALY-profiles well. If it does not, a preference for a more equal distribution of QALYs may be the product of two conceptually different factors: a preference for equality per se and diminishing marginal utility for QALYs. Diminishing marginal utility reflects that the policy maker's valuation of additional QALYs decreases with the amount of QALYs. For example, a policy maker may consider receiving 80 QALYs and receiving 90 QALYs as close because in both cases an individual has a long and healthy life, whereas he considers the difference between receiving 50 and 60 QALYs as larger. On the other hand, the policy maker might also prefer a more equal distribution regardless of his valuation of QALYs. Such a preference for equality is reflected in the equity weights π. The nonlinear rank-dependent QALY model allows to separate these two types of concern for equality and can, therefore, shed more light on what drives people's preferences over QALY-profiles.

As mentioned above, by taking a utility function over QALYs, our approach is consistent with Wagstaff (1991), Bleichrodt (1997) and Dolan (1998). In fact, Dolan's Cobb–Douglas model is a special case of Expression (2) in which the utility function over QALYs is logarithmic. Hence, our elicitation of the utility function over QALYs allows for a test of the Cobb–Douglas social welfare function proposed by Dolan.

Because the nonlinear rank-dependent QALY model, Expression (2), is more general than the rank-dependent QALY model, Expression (1), it shares the advantage of encompassing many of the social welfare functions that have been proposed in the literature. The nonlinear rank-dependent QALY model is also easy to use in practice once the utility function and the equity weighting function have been elicited. The elicitation of the nonlinear rank-dependent QALY model is, however, more involved because both the utility function and the equity weighting function must be assessed. We now turn to the issue of elicitation.

3. Elicitation

We elicited the nonlinear rank-dependent QALY model in two stages. In the first stage, the social utility function over QALYs was elicited. That is, we put subjects in the position of health policy makers and determined how they valued the amounts of QALYs received by others. This approach of putting subjects in the position of health policy makers is common in the literature on the equity-efficiency trade-off (e.g. Nord, 1993; Dolan, 1998; Rodrigues-Miguez and Pinto-Prades, 2002). The elicited social utilities were then used as inputs in the second stage, in which the equity weighting function was elicited.

3.1. Stage 1: Elicitation of the utility function

We first selected two gauge outcomes R and r and a starting value x_0. We took $x_0 > R > r$. Let (x, p, y) denote the rank-ordered QALY-profile that gives x QALYs to proportion p of the population and y QALYs to proportion $1-p$ of the population, $x \geq y$. We determined the number of QALYs x_1 that made a subject indifferent between (x_1, p, r) and (x_0, p, R). Because more QALYs are preferred to less, we must have $x_1 > x_0$. In terms of Expression (2), the indifference between (x_1, p, r) and (x_0, p, R) means that

$$w(p)U(x_1) + (1 - w(p))U(r) = w(p)U(x_0) + (1 - w(p))U(R) \tag{3a}$$

or

$$U(x_1) - U(x_0) = \frac{1 - w(p)}{w(p)}(U(R) - U(r)). \tag{3b}$$

After x_1 had been elicited, the number of QALYs x_2 was determined such that the subject was indifferent between (x_2, p, r) and (x_1, p, R). This indifference implies by Expression (2) that

$$U(x_2) - U(x_1) = \frac{1 - w(p)}{w(p)}(U(R)) - U(r)). \tag{4}$$

Combining Expressions (3b) and (4), we find that

$$U(x_2) - U(x_1) = U(x_1) - U(x_0) \tag{5}$$

We can continue in this fashion and elicit indifferences between (x_j, p, r) and (x_{j-1}, p, R), in the process eliciting a *standard sequence* $x_1, \ldots, x_k$ such that the utility intervals between successive elements are all equal. That is, $U(x_i) - U(x_{i-1}) = U(x_j) - U(x_{j-1})$ for all i and j between 1 and k.

The origin and the unit of the utility function can be chosen freely. We selected $U(x_0) = 0$ and $U(x_k) = 1$. It then follows that $U(x_j) = j/k$ for all j between 0 and k.

3.2. Stage 2: Elicitation of the equity weighting function

In the first stage of the elicitation procedure, the proportion p was kept constant to be able to elicit the utility function. To elicit the equity weighting function, the proportion p will be varied across questions. We used the following types of questions to elicit the equity weighting function. For low proportions p, we elicited the amount of QALYs z that made subjects indifferent between (x_k, p, x_0) and (x_i, p, z), $0 < i < k$, $x_i \geq z$, where the x's are elements of the standard sequence that was elicited in the first stage. Using the scaling $U(x_0) = 0$ and $U(x_k) = 1$, this indifference implies under the nonlinear rank-dependent QALY model:

$$\begin{aligned} & w(p)U(x_k) + (1 - w(p))U(x_0) = w(p)U(x_i) + (1 - w(p))U(z) \\ & \Leftrightarrow w(p) = w(p) \times (i/k) + (1 - w(p))U(z) \\ & \Leftrightarrow w(p) = \frac{U(z)}{1 + U(z) - (i/k)} \end{aligned} \tag{6a}$$

For high proportions p, we elicited indifference between (x_k, p, x_0) and (z, p, x_j), $0 < j < k$, $z \geq x_j$, where the x's, again, are elements of the standard sequence elicited in the first stage. By the nonlinear rank-dependent QALY model we obtain

$$\begin{aligned} & w(p)U(x_k) + (1 - w(p))U(x_0) = w(p)U(z) + (1 - w(p))U(x_j) \\ & \Leftrightarrow w(p) = w(p)U(z) + (1 - w(p)) \times (j/k) \\ & \Leftrightarrow w(p) = \frac{j/k}{1 + (j/k) - U(z)} \end{aligned} \tag{6b}$$

We could also have determined the equity weighting function by eliciting indifference between $(z, 1)$ and (x_k, p, x_0). This immediately gives $w(p) = U(z)$. Ubel et al. (2001) showed, however, that people tend to overstate their preference for equality when one of the options involves no inequality. We, therefore, avoided this type of questions.

Our procedure for determining the equity weights has three potential drawbacks. First, the outcomes z will generally not belong to the standard sequence elicited in the first stage and, therefore, their utility has to be approximated. This approximation may introduce bias. In the analysis of the results we used a linear approximation. Over small intervals the utility function does not deviate much from linearity and a linear approximation will be reasonable as long as successive elements of the standard sequence are close. To test the robustness of our results, we also approximated the utilities of z assuming three nonlinear parametric utilities, as will be described in Section 4.

Second, our procedure imposes bounds on the elicited equity weights. In Expression (6a), the equity weight can vary only between 0, which occurs when $z = x_0$ and i/k, which occurs when $z = x_i$. If the outcome z exceeds x_i then the QALY-profile (x_i, p, z) is no longer rank-ordered. Its rank-ordered analogue is $(z, 1 - p, x_i)$ and the indifference between (x_k, p, x_0) and $(z, 1 - p, x_i)$ gives by Expression (2)

$$w(p)U(x_k) + (1 - w(p))U(x_0) = w(1 - p)U(z) + (1 - w(1 - p))U(x_i) \tag{7}$$

That is, an equation with two unknowns, $w(p)$ and $w(1 - p)$, which cannot be solved in a unique manner. Similarly, in Expression (6b) the equity weight can only vary between j/k, which occurs when $z = x_j$, and 1, which occurs when $z = x_k$. In Section 4, we explain how we handled the potential boundedness problem.

Finally, our method may suffer from error propagation. Expressions (6a) and (6b) determine equity weights by a ratio. Error propagation for ratios can be problematic if the denominator is close to zero, so that small errors in the numerator lead to large errors in the ratio. Such problems do not occur in Expressions (6a) and (6b) because the denominator is far from zero, in fact more so than the numerator. Moreover, in both expressions, the numerator and the denominator are positively correlated because of a common term, which further reduces the overall error in the ratio. These observations suggest that error propagation will not be problematic in our design.

4. Experiments

4.1. Subjects

We performed two experiments to elicit the nonlinear rank-dependent QALY model. The subjects in the first experiment were 69 students at Erasmus University, Rotterdam. The subjects in the second experiment, which was run 1 month after the first, were 208 members from the general population. These subjects were recruited through a marketing agency from a representative sample of the Dutch population between 16 and 70 years old. Table 1 describes the characteristics of the sample from the general population split according to sex, level of education and age. Women were over-represented in our sample and people with a low level of education were slightly underrepresented.

Subjects in the general population sample were paid €17.50 for their participation, subjects in the students sample were paid €12.50. Prior to the actual experiments we performed nine pilot sessions, using students, to test and fine-tune the questionnaire.

Table 1
Characteristics of the sample from the general population

Population characteristics	Proportion
Sex	
Male	39.7
Female	60.3
Education	
Low	18.2
Middle	45.0
High	36.8
Age	
11–20	7.7
21–30	15.3
31–40	20.6
41–50	20.1
51–60	24.4
61–70	12.0

4.2. Procedures

The experiments consisted of a computer-based questionnaire. In the student sample, the experiment was carried out in personal interview sessions. In the general population sample, the experiment was carried out in group sessions with a maximum of 15 subjects per session. There were 22 group sessions in total and, hence, the average number of subjects per session was slightly less than 10. In the group sessions, the experiment was introduced classically. The questionnaire was then administered individually. There were three interviewers present during the group sessions to help subjects with any problems.

Before the experiment started we explained to the subjects why it is important for health policy to have information on people's preferences concerning the allocation of health and that their responses would help to make better-informed resource allocation decisions. We then explained to them in intuitive terms the concept of a QALY. The QALY-explanation that was read to the subjects can be found in Appendix A.

The decision problem in the actual experiment was the following. Subjects were asked to consider a cohort of newborns who suffer from some disease. The disease was left unspecified to avoid a possible framing effect. We deliberately selected a cohort of newborns to avoid that people thought they might themselves belong to the cohort and consider the decision problem as a decision under risk. In that case, preferences for equity would be confounded by risk attitude.

Subjects were told that there exist two treatments for the disease. The treatments have identical costs but differ in their effects. The treatments were labeled A and B to avoid possible framing effects. The outcomes of the treatments were integer numbers of QALYs. The treatments gave one part of the cohort, the "better-off group" as we will call them henceforth, more QALYs than the other part, the "worse-off group". Subjects were asked to make a choice between the two treatments. An example of the questions that subjects faced is given in Appendix B.

Following the explanation of the decision problem, subjects were given a practice question. In the student sample, we asked subjects to explain their answer to this question. In the general population sample, the interviewers asked some of the subjects to explain their answer. We used the explanation to check whether subjects understood the experimental task. In case we were convinced that subjects understood the task, we asked them to move on to the actual experiment.

Elicitation was by means of a sequence of choices. We opted for a choice-based elicitation procedure, because empirical evidence suggests that choice-based procedures are more consistent and less susceptible to biases than other elicitation procedures, such as matching (Bostic et al., 1990). We used the parameter estimation by sequential testing (PEST) procedure to elicit responses (Luce, 2000, pp. 291–292). PEST is an adaptive elicitation technique that determines the stimulus value for each new question by the subject's response to the previous one. PEST has the advantage of being able to home-in on an indifference value without the subject being aware that this is happening, thus, preventing the subject from forming a conscious numeric indifference. Such mental "matches" have been shown to lead to biases (see, Luce, 2000, for a review). Another advantage of the PEST procedure is that it tests for inconsistencies in subjects' responses, by repeating questions, and only converges to an indifference value when the responses become consistent. The PEST algo-

Table 2
Questions used to determine the equity weights

Proportion	Question	Interval
$p_1 = 1/6$	$(x_6, 1/6, 10)$ vs. $(x_2, 1/6, z_1)$	[0, 1/2]
$p_2 = 1/3$	$(x_6, 1/3, 10)$ vs. $(x_3, 1/3, z_2)$	[0, 2/3]
$p_3 = 1/2$	$(x_6, 1/2, 10)$ vs. $(x_4, 1/2, z_3)$	[0, 5/6]
$p_4 = 2/3$	$(x_6, 2/3, 10)$ vs. $(z_4, 2/3, x_2)$	[1/6, 1]
$p_5 = 5/6$	$(x_6, 5/6, 10)$ vs. $(z_5, 5/6, x_2)$	[1/6, 1]

rithm determined indifference to the nearest QALY integer value. The PEST algorithm is described in Appendix C, which also includes an illustration of the method.

In the first stage of the experiment, the utility function over QALYs was elicited. We elicited a standard sequence of six elements. So, $x_k = x_6$ in our study. The starting value x_0 was set equal to 10 QALYs and the two gauge outcomes R and r were set equal to 8 and 5 QALYs, respectively. We avoided the outcome 0 QALYs because this might invoke strong emotions which could distort the elicitation. The proportion p was set equal to 1/2. So, in the first stage of the experiment half of the cohort was in the better-off group and half was in the worse-off group and the outcome x_j was elicited so that indifference held between $(x_j, 1/2, 5)$ and $(x_{j-1}, 1/2, 8)$. We learned from the pilot sessions that these stimuli led to a standard sequence $x_1, \ldots, x_k$ whose successive elements were relatively close.

In the elicitation of the utility function, we varied only the outcome x_j to reach indifference. To try and avoid that subjects would focus too much on this outcome, and ignore the other stimuli, we included two filler questions in which all stimuli varied after each choice question.

To elicit the equity weighting function, the proportion of the cohort that belonged to the better-off group was varied. The elicitation of the equity weights was preceded by a practice question. By asking subjects to explain their answer to this question we were able to check whether they realized that the proportion had changed. We used five proportions in the elicitation of the equity weighting function: $p_1 = 1/6$, $p_2 = 1/3$, $p_3 = 1/2$, $p_4 = 2/3$ and $p_5 = 5/6$. The proportions were chosen so as to achieve a good spread over the [0, 1] interval and so that subjects could easily compute which treatment gave more QALYs. In the pilot sessions, we experimented with different proportions. It turned out that using smaller proportions than 1/6 or higher proportions than 5/6 led to unstable estimates. We, therefore, avoided using such low and high proportions in the actual experiment.

The stimuli varied with the proportion used. The first column of Table 2 shows the question that we employed for each proportion. In the questions for p_1, p_2 and p_3, we determined the outcome z that yielded indifference in the comparison between $(x_6, p, 10)$ and (x_i, p, z_m), $(i, m) = (2, 1), (3, 2), (4, 3)$, where x_i and x_6 were taken from the standard sequence that was elicited in the first stage, and we used Expression (6a) to compute the equity weights. If a subject was about to make a choice that implies that z_m exceeds x_i, in which case (x_i, p, z_m) is no longer rank-ordered and the analysis of Section 3 cannot be applied, the computer increased x_i to x_{i+1}. For example, in the question for p_1, x_2 was raised to x_3 when z_1 was about to exceed x_2. In the questions for p_4 and p_5, we elicited the outcome z_m that made the subject indifferent between $(x_6, p, 10)$ and (z_m, p, x_2), $m = 4, 5$, and we used Expression (6b) to compute the equity weights. In case a subject was about to violate rank-ordering of (z_m, p, x_2), which occurs if z_m is less than x_2, the computer decreased x_2 to x_1.

In Section 3, we explained that our elicitation method imposes bounds on the values that the equity weights can assume. The third column of Table 2 shows for each of the five proportions the interval within which the weight given to the better-off group is forced to lie. For example, the first entry of the column shows that the weight given to the better-off group when the size of the better-off group is 1/6 of the size of the cohort could never exceed 1/2. It would, of course, be better to have a higher upper bound than 1/2, which could be achieved by replacing x_2 by a "higher" element of the standard sequence, i.e. x_3, x_4 or x_5. We learned from the pilot sessions, however, that this made the estimates less stable and more sensitive to response error.

An example, using the data from one of our subjects, may explain the problem of sensitivity to response error. Let the standard sequence $\{x_1, \ldots, x_6\}$ be $\{15, 21, 29, 39, 54, 68\}$. Suppose that x_4 were used instead of x_2 to determine $w(1/6)$. We would then elicit the outcome z_1 that made the subject indifferent between (68, 1/6, 10) and (39, 1/6, z_1). Suppose that the subject's true equity weighting function is strictly convex, i.e. $w(p) < p$ for all p in (0, 1). To have $w(1/6) < 1/6$, z_1 must be smaller than 12. Suppose, as is likely, that the subject's choices are subject to response error. It is unlikely that a value of z_1 will be elicited that is lower than 11, because the subject will realize that by dominance (68, 1/6, 10) is better than (39, 1/6, 10). On the other hand, we may well elicit a value higher than 13. Hence, there is more room for errors "on the right" of 12 than "on the left". This asymmetric error pattern may bias $w(1/6)$ upwards. By using x_2, $w(1/6) < 1/6$ corresponds to a value of z of 14 or less. So, there is more room for error on the left and the problem of asymmetric error is less urgent. Only for those subjects who threatened to violate rank-dependency did the computer change x_2 into x_3. But the choices of these subjects implied $w(1/6) > 1/3$ and it is unlikely that these subjects' true value of $w(1/6)$ is less than 1/6, so the problem of asymmetric error did not occur for these subjects. To reduce the possibility of asymmetric errors affecting the results, we did not use proportions lower than 1/6 (or higher than 5/6) either. The final selection of the stimuli reflected what we believed to be the most finely tuned balance between stability of the estimates and restrictiveness of the bounds. In the pilot sessions, the bounds caused no problems: the implied equity weights were always at a safe distance from the bounds.

Because the PEST procedure requires a series of choices to find the indifference value, we were able to mix questions for different proportions. Both the outcomes and the proportions, therefore, changed across questions. We hoped that this would encourage subjects to focus on all the stimuli. The order in which questions appeared was random.

4.3. Analysis

We classified a subject's utility function as concave, linear or convex depending on how the slope of his elicited utility function changed across points of the standard sequence. Let ∇_{j-1}^{j} denote the difference between $(x_j - x_{j-1})$ and $(x_{j-1} - x_{j-2})$, $j = 2, \ldots, 6$. It is easily verified that a concave utility function corresponds to ∇_{j-1}^{j} positive, a linear utility function corresponds to ∇_{j-1}^{j} zero, and a convex utility function corresponds to ∇_{j-1}^{j} negative. We observed five values of ∇^j_{j-1} for each subject. To account for response error, we classified a subject's utility function as concave (linear/convex) if at least three values of ∇^j_{j-1} were positive (zero/negative).

To compute the equity weights, we needed the utilities of z_m, $m=1, \ldots, 5$. These were determined through linear interpolation. To test the robustness of the results, the utility of z_m, $m=1, \ldots, 5$, was also computed allowing for curvature of utility. We examined three parametric specifications for the utility function: the power function, the exponential function and the expo-power function.

Let $y=(x-x_0)/(x_6-x_0)$, where x is in $[x_0, x_6]$. The power function is defined by y^r, if $r>0$, by $\ln(y)$ if $r=0$, and by $-y^r$ if $r<0$. The exponential family is defined by $(e^{ry}-1)/(e^r-1)$ if $r\neq 0$ and by y if $r=0$. The power and exponential family are widely used in economics and (medical) decision analysis. Dolan's (1998) Cobb–Douglas social welfare function is the special case of Expression (2) where the utility function is logarithmic.

The expo-power family was introduced by Abdellaoui et al. (2002) and is a variation of a two-parameter family proposed by Saha (1993). The expo-power family is defined by $(1-\exp(-y^r/r))/(1-\exp(-1/r))$ with $r>0$. We included the expo-power family because it can accommodate some important preference patterns that are incompatible with both the power and the exponential family (see, Abdellaoui et al., 2002, for a discussion). The three utility functions were estimated by a distribution-free iterative procedure that minimized the sum of squared residuals, using the elements of the standard sequence and their corresponding utilities as data inputs.

To analyze equity weighting at the individual level, we examined how the slope of a subject's equity weighting function evolved. Let Δ^j_{j-1} be equal to the difference between $(w(p_j)-w(p_{j-1}))$ and $(w(p_{j-1})-w(p_{j-2}))$. For $j=2, \ldots, 6$, with $p_0=0$ and $p_6=1$, the function w is concave if Δ^j_{j-1} is positive for all j, linear if Δ^j_{j-1} is zero for all j and convex if Δ^j_{j-1} is negative for all j. Again, we allowed for response error in classifying subjects' weighting functions. The classification criterion used was motivated by a pattern observed in the data and will be explained in the next section.

5. Results

Four subjects had to be excluded from the student sample. Three of them did not reach convergence because they did not value additional QALYs above some level, one subject violated rank-ordering of QALY-profiles in the second stage even after the computer had adjusted the stimuli. This left 65 subjects in the analysis of the student sample.

In the general population sample, 29 subjects had to be excluded: 14 subjects violated rank-ordering of QALY-profiles even after adjustment of the stimuli, 9 subjects did not reach convergence because they did not value additional QALYs above some level, 4 subjects found the task too difficult, the computer of one subject crashed and 1 subject refused to start the experiment. This left 179 subjects in the analysis of the general population sample.

Whereas the other exclusions are unlikely to have affected the results, the exclusions due to violations of rank-ordering may have had an effect on the results. In the questions for

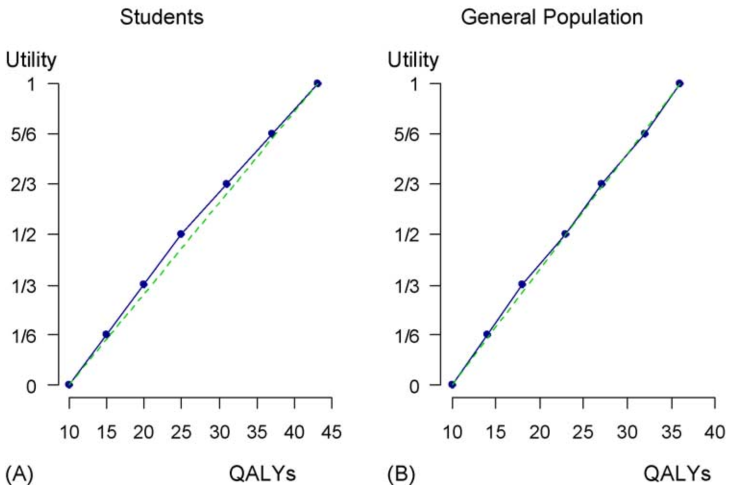


Fig. 1. The elicited utility functions.

p_1, p_2 or p_3, the violations of rank-ordering may have reflected a desire to make the profile (x_{i+1}, p, z_m) more attractive. If so, $w(p_i)$, $i = 1, 2, 3$, would have to exceed its imposed upper bound and the exclusion of these subjects leads to a downwards bias in the estimated equity weights. In the questions for p_4 and p_5, the violations of rank-ordering may have reflected a desire to make the profile (z_m, p, x_{j-1}) less attractive. If so, $w(p_j)$, $j = 4, 5$, would have to fall short of its imposed lower bound and the exclusion of these subjects leads to a downwards bias in the estimated equity weights.

Unfortunately, we do not know in which questions the subjects violated rank-ordering. We have some indication, however, that the effect of the violations was negligible, as we discuss in Section 7.

5.1. Elicitation of the utility function

Fig. 1 shows the elicited utility functions over QALYs for both samples, based on the median data.[1] Both utility functions were close to linear: the utility function for the student sample was slightly concave, whereas the utility function for the general population sample might be described as "linear with random error".

The above observations were confirmed when we looked at the parametric estimates of the utility function, which are displayed in Table 3. The linear utility function is the special case of the power function when the power coefficient is equal to 1, and it is the special case of the exponential function when the exponent is equal to 0. Table 3 shows that in both samples, the mean and median power parameters were close to 1 and the mean and median exponential parameters were close to 0, suggesting that the assumption of linear utility over

[1] The functions look similar when we use the mean data.

Table 3
Parameter estimates

	Parametric families								
	Power			Exponential			Expo-power		
	Median	Mean	IQR	Median	Mean	IQR	Median	Mean	IQR
Students	0.90	0.97	0.21	−0.32	−0.21	0.67	1.18	1.25	0.22
General population	0.96	1.07	0.32	−0.06	−0.01	0.98	1.25	1.35	0.34

Note: IQR stands for inter-quartile range.

QALYs was reasonable at the aggregate level. The inter-quartile ranges show that individual coefficients varied considerably and that the above conclusion did not necessarily hold at the individual level.

The power coefficient was significantly different from zero, the case where utility is logarithmic, suggesting that Dolan's (1998) Cobb–Douglas social welfare function did not fit our data well. For all three estimations, no significant differences were found between the coefficients in the student sample and those in the general population sample. We found in neither sample a significant difference in goodness of fit between the three parametric specifications.

Table 4, which displays the results of the analysis of the individual data, shows that that there was no predominant shape of the social utility function. In both samples, the proportion of subjects with a concave utility function, which corresponds to diminishing marginal utility, was slightly higher than either of the two other categories, but not significantly so. One reason why there were relatively many subjects whose utility function could not be classified is that we used a rather strict classification criterion. For example, if a subject's standard sequence was equal to $\{10, 15, 19, 24, 28, 33\}$ then he was classified as mixed even though his utility function was almost perfectly linear. We could of course have used a weaker classification criterion, e.g. the sign of ∇_{j-1}^{j} plus or minus the standard deviation of the responses, but this would have allowed for the possibility that, for a given ∇_{j-1}^{j}, a subject's utility function was both classified as convex and as concave, which seemed undesirable.

An easy heuristic for subjects to use in answering the utility elicitation questions would be to let $x_j - x_{j-1} = R - r$. This might have inflated support for the linear utility function. There were three subjects in the student sample and one in the general population sample who had such an answer pattern. This suggests that a large majority of our subjects did not use such a heuristic.

Table 4
Classification of subjects in terms of the shape of their utility function

	Concave (%)	Linear (%)	Convex (%)
Students	21.5	18.5	12.3
General population	20.7	17.9	17.3

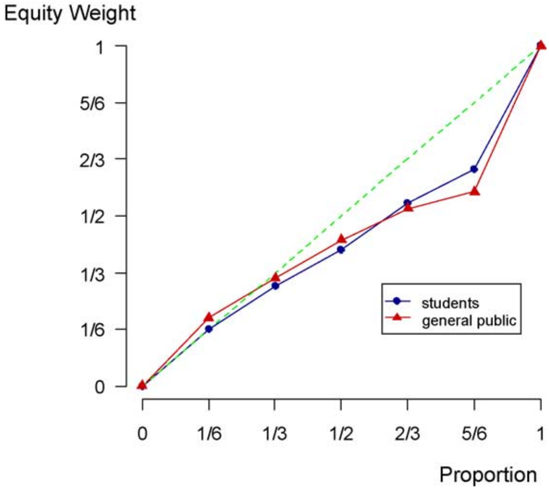


Fig. 2. The elicited equity weighting functions.

5.2. *Elicitation of the equity weighting function*

The number (proportion) of cases in which the computer had to adjust the stimuli to avoid a violation of rank-ordering was 15 (21.7%), 7 (10.1%), 0, 5 (7.2%) and 3 (4.3%) in questions 1–5 in the student sample and 75 (41.9%), 35 (19.6%), 15 (8.4%), 38 (21.2%) and 29 (16.2%) in the general population sample.

The conclusions did not depend on whether we used linear utility, power utility, exponential utility, or expo-power utility to compute the utilities of the z_m, $m = 1, \ldots, 5$. We, therefore, only report the results under the linear approximation.

Fig. 2 shows the median equity weighting functions for both samples. The shape was similar: it was largely convex except for the first part which was linear for the student sample and slightly concave for the general population sample. Recall from Section 2 that a convex (linear/concave) weighting function corresponds to inequality aversion (neutrality/seeking). Fig. 2, therefore, suggests that subjects were predominantly averse to inequalities in health, except when the size of the better-off group was small.

One possible reason why subjects may not have been uniformly inequality averse is that they did not properly take into account group size. There is a vast psychological literature showing that when people are dealing with relative frequencies, like proportions, they distort them in recognizable ways: people tend to overestimate small proportions and underestimate high proportions (e.g. Tversky and Kahneman, 1992; Gonzalez and Wu, 1999). We will label this type of behavior *insensitivity to group size*. In our study, insensitivity to group size would make that people perceive the better-off group as larger than it actually is when the proportion in the better-off group is small, say 1/6, and perceive the better-off group as smaller than it actually is when the proportion in the better-off group is high, say 5/6.

Table 5
Classification of subjects in terms of the shape of their equity weighting function

	Concave (%)	Linear (%)	Convex (%)	Insensitivity (%)
Students	7.7	0	41.5	38.5
General population	3.9	0	31.3	54.2

Table 5 shows the results of the individual analyses of the equity weighting functions. The hypothesis of insensitivity to group size, formulated above, would support concavity of the equity weighting function when the proportion of the better-off group is close to 0 and convexity of the equity weighting function when the proportion is close to 1. In other words, insensitivity to group size is inconsistent with convexity of the equity weighting function when the proportion of the better-off group is close to 0, and is inconsistent with concavity of the equity weighting function when this proportion is close to 1. To account for both insensitivity to group size and response error, we classified a subject's equity weighting function as concave if at least three values of Δ^j_{j-1} were positive and $(1 - w(5/6)) \leq 1/6$, i.e. there was no "downwards jump" in the equity weights near 1, as convex if at least three values of Δ^j_{j-1} were negative and $w(1/6) \leq 1/6$, i.e. there was no "upwards jump" in the equity weights near 0, and as linear if at least three values of Δ^j_{j-1} were zero and not both $w(1/6) > 1/6$ and $(1 - w(5/6)) > 1/6$.[2]

Table 5 shows that few subjects had a concave or linear equity weighting function. The proportion of subjects with a convex equity weighting function was much higher although still lower than 50%. The final column of Table 5, which shows the proportion of subjects for whom both $w(1/6) > 1/6$ and $(1 - w(5/6)) > 1/6$, suggests that the behavior of a sizeable number of our subjects was consistent with insensitivity to group size.

The above analysis suggests that the equity weights that we obtained were the product of both insensitivity to group size and what we may call "true" concerns for equality. To try and separate these two factors, we estimated the following parametric form for the equity weighting function:

$$w(p) = \frac{\delta p^{\gamma}}{\delta p^{\gamma} + (1 - p)^{\gamma}} \tag{8}$$

This specification was first proposed by Goldstein and Einhorn (1987) for decision under risk. Gonzalez and Wu (1999) gave an interpretation for the parameters γ and δ, which with some modifications, also applies to the social decision context that we consider here. The parameter γ determines the curvature of $w(p)$ and, hence, the sensitivity to group size. Values less than 1 indicate insensitivity to group size and the lower γ is, the less sensitive the individual is to changes in group size. The parameter δ indicates the attractiveness of giving health gains to the better-off group, and thus measures preferences for equality. The

[2] To test the robustness of our findings, we also used the classification concave if at least three values of Δ^j_{j-1} were positive and Δ^6_5 was not negative, linear if at least three values of Δ^j_{j-1} were zero and not both Δ^6_5 negative and Δ^2_1 positive and convex if at least three values of Δ^j_{j-1} were negative and Δ^2_1 was not positive. The results were similar to those reported in Table 5.

Table 6
Parameter estimates for the equity weighting function (Expression (8))

	Parameters					
	Gamma			Delta		
	Median	Mean	IQR	Median	Mean	IQR
Students	0.68	0.69	0.32	0.59	0.69	0.32
General population	0.56	0.61	0.44	0.63	0.73	0.47

Note: IQR stands for inter-quartile range.

lower δ is, the more equality-minded people are. Values less than 1 correspond to inequality aversion.

Expression (8) was estimated by a distribution-free iterative procedure that minimized the sum of the squared residuals. Table 6, which displays the results of the estimation, shows that insensitivity to group size and preferences for equality jointly determined the equity weights. Insensitivity to group size was stronger in the general population sample. The difference in the estimate of γ is significant by the nonparametric Mann–Whitney test ($P = 0.004$), but only marginally so by the independent samples t-test ($P = 0.067$). Aversion to inequality, measured by δ, was similar in the two samples.

Because the parameters γ and δ are largely independent, we can use Expression (8) to correct for the impact of insensitivity to group size on the equity weights by setting $\gamma = 1$. Fig. 3 shows the equity weights when $\gamma = 1$ and $\delta = 0.6$, the case which corresponds to our median data. The figure shows, for example, that the weight given to the better-off group was 0.375 when the size of the better-off group was equal to half the cohort.

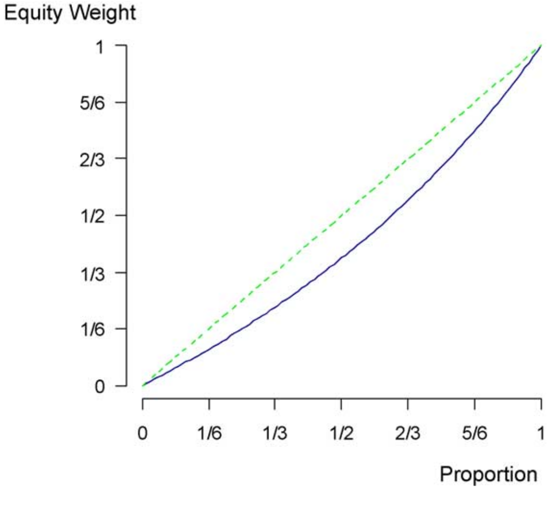


Fig. 3. The elicited equity weighting function after correction for insensitivity to group size.

6. Implementation in health policy

To illustrate the implications of our findings, we computed equity-adjusted cost-utility ratios for 12 treatments. To perform these computations, we made two assumptions. These assumptions are not innocuous and we therefore urge the reader to interpret the equity-adjusted cost-utility ratios with caution. These ratios serve as an illustration of how our method can be applied in practice, not as a guide to policy making. The first assumption is that we can extrapolate outside the domain of estimations. We found that the social utility function over QALYs was roughly linear on the interval [10, 40]. The data suggest that linearity also held on [5, 10]. Linearity on [10, 40] means that $U(x_j) - U(x_{j-1})$, $j = 1, \ldots, 6$, is about 5. From Expressions (4) and (5), we know that $U(x_j) - U(x_{j-1})$ is equal to $((1 - w(p))/w(p))(U(8) - U(5))$. We also found that $w(1/2)$ was about equal to 0.4. This implies that $U(8) - U(5)$ was close to 3, which is consistent with linearity. However, we do not know whether U was also linear on [0, 5] and on [40, →). In fact, when we looked at those subjects for whom x_6 exceeded 50 years then we found more concavity than in the general sample, suggesting that the assumption of linearity of the social utility function on [40, →) need not hold. Similarly, we did not estimate any equity weights on (0, 1/6) and on (5/6, 1). We had to assume that the estimated pattern of equity weighting on [1/6, 5/6] can be extrapolated to these two subdomains.

Our second assumption is that it is better to use the equity weights that are corrected for insensitivity to group size than the uncorrected ones. That is, we will use Expression (8) with $\gamma = 1$. We used the corrected equity weights because we believe that insensitivity to group size, which arises because of people's limited cognitive abilities, is a bias in people's preferences that ought to have no impact on health policy. We realize that this assumption is controversial. After all, some of what we are correcting for may be true equity preference. Nevertheless, we believe that the corrected equity weights were closer to subjects' true equity weights than the uncorrected weights.

Most of the selected treatments were taken from Stolk et al. (2003), data on the remaining conditions were obtained through personal communication. To adjust cost-utility ratios for equity considerations we computed the distribution of QALYs within the Netherlands, on the basis of mortality figures (CBS, 2003) and quality of life estimates (Toenders, 2002). The distribution is displayed in the first two columns of Table 7. We then computed the equity weights for a patient in each of the groups. The equity weights were computed using Expression (8). The third column shows the equity weights when we used the parameter values that best fitted our data in the general population sample, $\gamma = 0.56$ and $\delta = 0.63$, the fourth column shows the equity weights after correction for insensitivity to group size, $\gamma = 1$, and using $\delta = 0.60$. We rescaled the equity weights so that the weight given to a patient with expected lifetime QALYs between 65 and 70 was equal to 1. This scaling is based on Williams (1997), who suggested that a person's fair innings was approximately 70 QALYs. The third column of the table shows the effect of insensitivity to group size: individuals who are in the tails of the QALY distribution get more weight than those who are closer to the middle of the distribution. The fourth column shows that this, counterintuitive, effect disappears after correction for insensitivity to group size. Then, the weights are monotonically decreasing.

Table 7
Distribution of QALYs and equity weights

Lifetime QALYs	Proportion	Equity weight	
		$\gamma = 0.56$, $\delta = 0.63$	$\gamma = 1$, $\delta = 0.6$
<1	0.55	26.81	1.56
1–15	0.27	12.29	1.55
15–30	0.73	8.87	1.54
30–40	1.06	6.18	1.52
40–50	3.15	4.04	1.48
50–55	4.28	2.66	1.41
55–60	6.40	1.88	1.32
60–65	11.07	1.36	1.19
65–70	20.38	1	1
70–75	26.54	0.88	0.79
75–80	21.40	1.24	0.64
80–82.5	3.32	2.82	0.57
>82.5	0.85	8.23	0.56

Table 8 displays the results of adjusting the cost-utility ratios for equity concerns. The first column describes the conditions that we studied, the second the treatments for these conditions. The third column shows for each treatment the costs per QALY gained when no equity weighting was applied, i.e. under the common procedure of aggregating QALYs. The fourth column shows the ranking of the treatments in terms of cost-effectiveness when

Table 8
Equity-adjusted cost-utility ratios

Condition	Treatment	Cost/ QALY	Rank	Lifetime QALYs	Equity weight ($\gamma = 1$, $\delta = 0.6$)	Equity-adjusted cost/QALY	Rank
Congenital anorectal malformation	Surgery	2482	1	9.4	1.55	1601	1
Erectile dysfunction	Sildenafil	5656	2	77.0	0.64	8838	3
Non-Hodgkin lymphoma	Chemotherapy	7771	3	73.3	0.79	9837	4
Artherosclerosis	Clopidogrel	11629	4	54.9	1.41	8248	2
Benign prostatic obstruction	Finasteride	12788	5	80.7	0.57	22435	7
Onychomycosis	Terbinafine	16843	6	83.7	0.57	29549	9
Osteoporosis	Oestrogen	18151	7	83.2	0.57	31844	11
High cholesterol	Statins	18151	7	56.1	1.41	12873	5
Metastatic breast cancer	Chemotherapy	22441	9	56.1	1.41	15916	6
Heart disease	Heart transplant	38206	10	42.2	1.48	25815	8
End-stage renal disease	Kidney replacement	44607	11	57.8	1.41	31636	10
Pulmonary hypertension	Lung transplant	79412	12	41.6	1.48	53657	12

no equity weighting was applied. As the table shows, surgery for congenital anorectal malformation was the most cost-effective treatment and lung transplantation for pulmonary hypertension was the least cost-effective treatment.

The fifth column shows for each disease the number of expected lifetime QALYs that the average patient obtains without treatment. The sixth column gives the rescaled equity weights that obtained after correction for insensitivity to group size. These weights can directly be read off from Table 7. The seventh column shows the cost-utility ratios adjusted by these equity weights. The final column shows the ranking of the treatments in terms of equity-adjusted cost-utility ratios. As expected, there were some shifts in ranking in favor of treatments aimed at patients with lower expected lifetime QALYs. For example, the cost per QALY of statins was higher than that of terbinafine when no equity weighting was applied, but statins were more cost-effective than terbinafine when cost-utility ratios were adjusted for equity concerns.

7. Discussion

7.1. Main findings

In this paper, we have elicited, both in a sample of students and in a sample from the general population, the trade-off between equity and efficiency in the allocation of health. We assumed the nonlinear rank-dependent QALY model, a model that encompasses many of the social welfare functions that have been proposed in the literature. A correction for utility curvature was applied but we found that, on the aggregate level, social preferences were approximately linear in QALYs. People were generally inequality averse, except when the better-off group was small. The reason why we found no global inequality aversion may be insensitivity to group size. Global inequality aversion was observed when we corrected for insensitivity to group size. Few differences were observed between the sample of students and the sample from the general population.

7.2. Possible biases

As noted in Section 5, the exclusions due to violations of rank-dependency may have affected the results. We tested for the effect of these exclusions by making the extreme assumption that the excluded subjects violated rank-dependency in every question. This assumption means that these subjects had the highest equity weights of all subjects in the questions for p_1, p_2 and p_3, and the lowest equity weights in the questions for p_4 and p_5. Such a preference pattern is unlikely and the assumption is almost certainly too extreme, which means that the actual bias will be smaller, but the analysis gives an indication of the maximum effect of the exclusions on the median equity weights. Under the assumption, the median equity weights for 1/6, 1/3, 1/2, 2/3 and 5/6 changed by 0, +0.003, +0.006, −0.007 and 0, respectively, in the student sample and by 0, +0.015, +0.026, −0.022, −0.025 in the general population sample. Hence, even under an extreme assumption about the effect of the exclusions due to violations of rank-ordering, the effect of these exclusions was small.

A frequently encountered problem in preference assessment tasks is that people have a tendency to respond in round numbers, often multiples of five, which can lead to bias. Because a choice-based procedure was used, round answers were less likely in our study. In fact, the proportion of round answers (multiples of five) was 21.7% in the student sample and 21.3% in the general population sample, which are not significantly different from 20%, the proportion of round answers expected when people do not have a tendency to use round answers.

It may have been possible that some subjects did not understand the concept of a QALY properly, leading to additional response error. It would have been easier to perform the experiment with years of life instead of QALYs. We opted to use QALYs, because policy makers and researchers are most interested in the trade-off between equity and efficiency as measured by QALYs. Upon questioning by the experimenter, most subjects seemed to understand the concept of a QALY well. To complete the exercise, they generally assumed that people in the cohort lived in relatively good health for the largest part of their life, and that the largest QALY loss was related to life-years lost.

Our findings depend on the validity of the nonlinear rank-dependent QALY model, Expression (2). Even though Expression (2) is quite general, it may in some cases be too restrictive. The model assumes, in particular, that the equity weights depend only on individuals' relative positions, their rank, and not on absolute differences between the amounts of QALYs received. If this assumption does not hold then our results may no longer be valid. Another violation would occur if there is no separability between the equity weights and the utility for QALYs. In that case, the elicitation of the utility for QALYs might depend on the proportion used. We could have used a more general model than Expression (2) to take these possible violations into account. This would, however, have led to a model that is more difficult to apply in practice. The question is whether violations of the nonlinear rank-dependent QALY model, if any, are sufficiently widespread and serious to justify giving up the tractability of the model.

Finally, it is possible that, even though we tried to control for it, asymmetric errors may still have affected the results. If this were true, then these errors will have had most effect on $w(1/6)$ and $w(5/6)$, biasing $w(1/6)$ upwards and $w(5/6)$ downwards. The effect on the other three weights that we elicited is probably negligible, because in these estimations the stimuli were not close to the bounds and there was enough room for error "on both sides". Our main finding of a generally convex equity weighting function, i.e. aversion to inequality, is confirmed when we only look at $w(1/3)$, $w(1/2)$ and $w(2/3)$, giving grounds for confidence in the results.

7.3. Final remarks

Our study suggests that people are averse to inequalities in health. If people's societal preferences ought to have a place in health policy, then our findings connote that QALYs should be weighted for equity concerns. We have shown that the rank-dependent QALY model can be used for this: we have presented a method to elicit the equity weights under the model and we have shown how these equity weights can be implemented in health policy. We repeat that the purpose of the latter exercise was illustrative; before more robustness checks are performed, restraint should be exercised in using the data we presented in actual policy making.

Finally, a few words about the equity concept we used are in order. Because we studied people's preferences over allocations of lifetime QALYs, our study focused on differences in lifetime health expectancy between groups of newborns. This setup implicitly assumed that the desirability of a distribution depends on people's (expected) lifetime health. In that sense, our approach is close to Williams' fair innings approach. Several authors have discussed other concepts of equity and have argued that equity may also be concerned with other issues, such as patients' actual health state and when and how health losses occur (Culyer and Wagstaff, 1993; Cuadras et al., 2001; Dolan and Olsen, 2001). Our empirical results have little bearing in case such equity concerns are adopted. How these other equity concerns can be operationalized, remains, therefore, an open question.

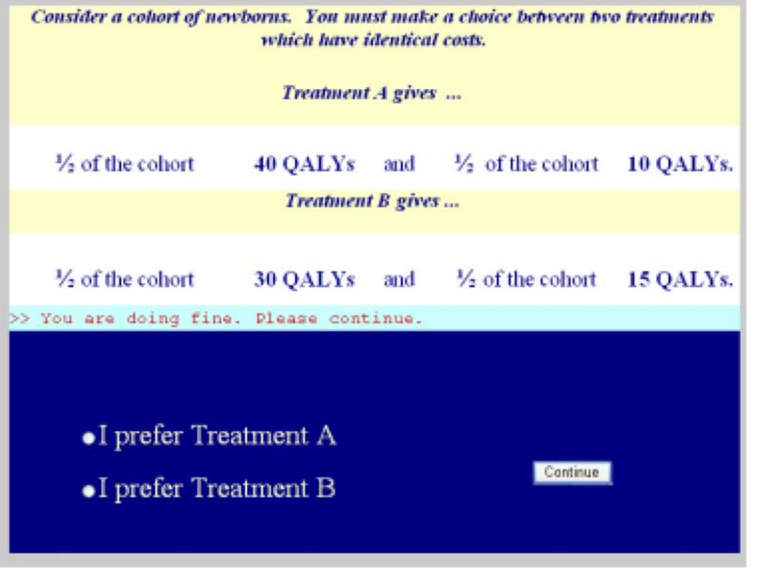
Acknowledgements

Peter Wakker, two anonymous reviewers and participants of the Workshop on QALYs in Alicante gave many helpful comments. We acknowledge financial support from the Netherlands Organisation for Scientific Research (NWO) under Project DO 945-10-034 and from the Netherlands Organisation for Health Research and Development (ZON-MW) under Project VVS-2-7. Han Bleichrodt also acknowledges a VIDI-Grant from NWO.

Appendix A. Explanation of QALYs

In this experiment, health is described in terms of quality-adjusted life-years (QALYs). Quality-adjusted life-years are a measure of health and can be calculated by multiplying life-years by a numeric value that reflects quality of life during those years. A year in full health counts as 1 QALY. A year in which people are confronted with health problems counts as less than 1 QALY. For example, I consider myself to be in full health. As long as I stay in full health each year I live counts as 1 QALY. But suppose that I had arthritis then each year would count as less than 1 QALY. If we assume, for example, that pain and mobility reduce my quality of life by 50%, then each year that I live in this health state counts as 1/2 QALY. The questionnaire specifies how many QALYs a subgroup of a cohort will get. If the number of QALYs is high, you can be sure that the people live long and that their quality of life is good. If the number of QALYs is low, then this number of QALYs can be the result of either a long life with severe disability or a short life with no disability.

Appendix B. Presentation of the experimental questions



Appendix C. Explanation of the PEST procedure

The PEST procedure obeys the following four rules:

1. On every reversal of step direction, halve the step size.
2. The second step in a given direction, if called for, is the same size as the first.
3. The fourth and subsequent steps in a given direction are each double their predecessor, except that large steps may be disturbing to a human observer and an upper limit on permissible step size may be needed.
4. Whether a third successive step in a given direction is the same as or double the second depends on the sequence of steps leading to the most recent reversal. If the step immediately preceding that reversal resulted from a doubling, then the third step is not doubled, while if the step leading to the most recent reversal was not the result of a doubling, then this third step is double the second. Doubling occurs on the first three responses in the same direction

Consider the following example:

A. 1/2 the cohort gets *X* QALYS and 1/2 the cohort gets 5 QALYs;
B. 1/2 cohort gets 30 QALYs and 1/2 the cohort gets 8 QALYs.

The initial increment for change was 4 QALYs. The stopping rule occurred when an increment change in QALYs in option A is less than 2 QALYs. The first step is to select a

Table A.1
Illustration of the PEST procedure

Trial	*X*	Choice	Comment
1	70	A	Random selection
2	66	A	First change
3	62	A	Rule 2
4	54	A	Rule 4
5	38	B	Rule 3
6	46	A	Rule 1
7	42	B	Rule 1
8	40	A	Rule 1
9	41	B	Stopping rule

random starting value of X in some interval, say (30, 100). This interval depended on the stimuli in the question. Suppose that $X = 70$. Table A.1 illustrates the PEST procedure.

Note that the PEST procedure can correct for errors. In the example above we began zeroing in at Trial 5. However, if a subject got to Trial 7 and had made some errors, he could break out of the convergence by choosing A A A or B B B during the next several trials. As mentioned in the main text, we included random ‘filler’ trials at a ratio of 2 random to every 1 real trial so that the subject did not know that convergence was happening.

References

Abdellaoui, M., Barrios C., Wakker P.P., 2002. Reconciling introspective utility with revealed preference: experimental arguments based on prospect theory. Working Paper, ENSAM/ESTP, Cachan.

Abellan, J.M., Pinto, J.L., 1999. Health state after treatment: a reason for discrimination? Health Economics 8, 701–707.

Atkinson, A.B., 1970. On the measurement of inequality. Journal of Economic Theory 2, 244–263.

Bleichrodt, H., 1997. Health utility indices and equity considerations. Journal of Health Economics 16, 65–91.

Bleichrodt, H., Diecidue, E., Quiggin, J., 2004. Equity weights in the allocation of health care: the rank-dependent QALY model. Journal of Health Economics 23, 157–171.

Bostic, R., Herrnstein, R.J., Luce, R.D., 1990. The effect on the preference reversal of using choice indifferences. Journal of Economic Behavior and Organization 13, 193–212.

CBS, 2003. Statline.

Cuadras, X., Pinto, J.L., Abellan, J.M., 2001. Equity considerations in health care: the relevance of claims. Health Economics 10, 187–205.

Culyer, A.J., Wagstaff, A., 1993. Equity and equality in health care. Journal of Health Economics 12, 431–457.

Dolan, P., 1998. The measurement of individual utility and social welfare. Journal of Health Economics 17, 39–52.

Dolan, P., Olsen, J.A., 2001. Equity in health: the importance of different health streams. Journal of Health Economics 20, 823–834.

Ebert, U., 1988. Measurement of inequality: an attempt at unification and generalization. Social Choice and Welfare 5, 147–169.

Goldstein, W.M., Einhorn, H.J., 1987. Expression theory and the preference reversal phenomena. Psychological Review 94, 236–254.

Gonzalez, R., Wu, G., 1999. On the form of the probability weighting function. Cognitive Psychology 38, 129–166.

Harris, J., 1987. Qalyfying the value of life. Journal of Medical Ethics 13, 117–123.

Luce, R.D., 2000. Utility of Gains and Losses: Measurement-Theoretical and Experimental Approaches. Lawrence Erlbaum Associates, Inc., Mahwah, New Jersey.

Nord, E., 1993. The trade-off between severity of illness and treatment effect in cost-value analysis of health care. Health Policy 24, 227–238.

Nord, E., 1995. The person-trade-off approach to valuing health care programs. Medical Decision Making 15, 201–208.

Rodrigues-Miguez, E., Pinto-Prades, J.L., 2002. Measuring the social importance of concentration or dispersion of individual health benefits. Health Economics 11, 43–53.

Saha, A., 1993. Expo-power utility: a 'flexible' form for absolute and relative risk aversion. American Journal of Agricultural Economics 75, 905–913.

Stolk, E.A., Van Donselaar G., Brouwer W.B.F., Van Busschbach J.J., 2003. Defining equity in the equity-efficiency tradeoff: an exploration of the concept of proportional shortfall. Working Paper, Erasmus University.

Toenders, W.G.M., 2002. Vervolgonderzoek breedte geneesmiddelenpakket. College voor Zorgverzekeringen, Amstelveen.

Tversky, A., Kahneman, D., 1992. Advances in prospect theory: cumulative representation of uncertainty. Journal of Risk and Uncertainty 5, 297–323.

Ubel, P.A., Baron, J., Asch, D.A., 2001. Preference for equity as a framing effect. Medical Decision Making 21, 180–189.

Wagstaff, A., 1991. QALYs and the equity-efficiency trade-off. Journal of Health Economics 10, 21–41.

Wagstaff, A., 1993. Erratum. Journal of Health Economics 12, 237.

Wakker, P.P., Deneffe, D., 1996. Eliciting von Neumann-Morgenstern utilities when probabilities are distorted or unknown. Management Science 42, 1131–1150.

Williams, A., 1997. Intergenerational equity: an exploration of the 'fair innings' argument. Health Economics 6, 117–132.

Williams, A., Cookson, R., 2000. Equity in health. In: Culyer, A.J., Newhouse, J.P. (Eds.), Handbook of Health Economics. Elsevier, Amsterdam, pp. 1863–1910.