THE UNIVERSITY of York

CENTRE FOR HEALTH ECONOMICS

The Value of Health at Different Ages

Aki Tsuchiya

DISCUSSION PAPER 184

The Value of Health at Different Ages

Aki Tsuchiya

February 2001

ABSTRACT

Is the value of being healthy the same across all ages? The standard practice of economic evaluation assumes so, and does not discriminate between a QALY (Quality Adjusted Life Year) to an elderly person and one to a child. But on the other hand, it is possible to assume that the value of a healthy year is different according to age, as has been done with DALYs (Disability Adjusted Life Years).

This paper is based on a series of interviews designed to elicit and to quantify preferences concerning health at different ages. There were three hypotheses to be tested: (1) that the relative value of health decreases with age, (2) that this decreasing profile is independent of a respondent's age, and (3) that this age-related preference can be expressed on an interval scale. The results obtained did turn out to depend on a respondent's age: a mostly negative age-value profile was obtained from younger respondents, but the profile from older respondents had a peak at middle age. Thus, the 1st and 2nd hypotheses were largely rejected. The 3rd hypothesis cannot be rejected, but it should be noted that the variance of the responses was large, thus rendering rejection somewhat less likely.

To conclude, the respondents valued a unit of health differently, depending on the age of the patient. While this study does not attempt to determine the exact continuous age-value profile, it found the profile clearly declining beyond middle age.

1. INTRODUCTION

Let us think of the social value of being healthy at different ages. Is the social value of being healthy for one year at age five equal to the value of being healthy for one year at age 60? The established practice in the paradigm of QALYs (Quality Adjusted Life Years) today assumes that an additional QALY is of the same value no matter to whom it accrues. This implies that the social value of a QALY, and hence of health, is invariant with age.[1] It may be possible to refer to this practice as "egalitarian" and view it as an essential requirement for the ethical acceptability of QALYs. Nevertheless, there are three things to note. Firstly, it is technically feasible to "age weight" QALYs: for example, DALYs (Disability Adjusted Life Years) is an example which incorporates such age weighting (Murray and Lopez,1994; Murray,1996; World Bank,1993). Secondly, on the theoretical front, it is possible to argue for age weighting based on equity concerns: the "fair innings argument" supports valuing health and life according to people's age (Williams,1997; Tsuchiya,2000). Thirdly, on the empirical level, there is some evidence that the general public does not evaluate health benefits independently of patient age. Tsuchiya (1999) surveys several empirical studies of age-related preferences amongst the general public, and finds that treating QALYs equally irrespectively of patient age is more often rejected than supported.

Busschbach et al.(1993; hereafter referred to as the BHC study) is a study that is based on a series of interviews that elicited the extent of preferences[2] regarding health improvements at different ages. The importance of this study lies in the way in which the value of a given health improvement at different age was directly elicited, and not indirectly through the value of remaining life at different ages. The results of the BHC study were: (1) the value of health at infancy has approximately twice the value of health at advanced ages, with middle age in-between; (2) two different respondent groups (an elderly group and a student group) were used, and not only the student group but the elderly group themselves support this declining profile; and (3) the preferences can be expressed on interval scales.

This paper concerns a (quasi-)replication of the interviews as conducted in the above study, and reports on the findings. Further, there are some additional analyses focusing on the individual responses.

2. METHODS

2.1 The interview

The interviews carried out in Japan followed the BHC interviews as closely as possible. The latter were designed to compare the value of health at different ages, perceived by convenience samples of the general public. Respondents were asked to prioritise between two patients competing for limited resources, where the only known attribute of the two patients was their age: other factors such as biological/physiological conditions, family background and occupation etc. were assumed to be either equivalent or unknown.

While there have been several empirical studies of preferences concerning health at different ages, the main characteristic of the BHC study is that it has quantified the magnitude of this preference, employing a hybrid of time trade-off (TTO) and the so-called person trade-off (PTO). Amongst the available methodologies to elicit and quantify people's relative preferences for different health states, PTO has been re-evaluated recently (Nord,1995). This method was known in the 1970s under the name Equivalence of Numbers (Patrick et al.,1973), but as Standard Gamble (SG) and TTO became more and more popular, it has long been left behind. While SG and TTO, in their conventional formulation, typically ask respondents to "imagine *yourself* in a certain health state", and are thus designed to elicit personal preferences of the individual respondents,

[1] It is also implied that the social value of a QALY is invariant with sex, social class, income, willingness (and ability) to pay, ethnicity, etc.

[2] As the title shows, the BHC study refers to the "utility" of health. However, since the interview questions ask about the respondent's preferences concerning priority between imaginary patients other than the respondents themselves, and not about their own utility, this paper will refer to the concept as "preference".

PTO asks for judgements of the respondents at a social level, by asking for example to “imagine that there are a certain number of people, *not yourself*, in a certain health state”. Although the BHC paper itself does not discuss these methodological issues, the method employed is *a TTO from this societal point of view*, and as such, while the trade-off is in terms of time, it involves an intra-personal perspective of the PTO kind.

In the original interviews and in the replication, the respondents were asked to imagine themselves as a director of a hospital, and to consider a situation where there were two patients but only one of them could be fully treated without delay, and the other patient had to wait for a certain period of time. The first stage of the scenario was set as follows: the two patients have kidney failure; there is only one transplantable kidney; the patient that does not receive the organ now will have to wait for two years on dialysis; and the only relevant difference between them is their age. Furthermore, there were four assumptions:

(1) both dialysis and organ transplantation are to be completely riskless,

(2) the transplanted patient will regain perfect health; ie there are to be no complications or side-effects related to transplantation,

(3) the patient not operated now will wait for two years on dialysis; a transplantable kidney then will be available for certain; and the patient then will have the priority to receive it, and

(4) no matter at what point in time the organ transplant is administered, the two patients will live for certain to the same age.

For instance, when asked to choose between a 35-year-old patient and a five-year-old patient, suppose a respondent chose the 35-year-old patient. Then, for the second stage of the question, there would be a slight re-arrangement of the conditions. Now the preferred patient would have to wait for one year if not operated today, instead of the initial two years; the waiting time for the unselected patient would remain at two years throughout the exercise. In other words, would the respondent still prefer to choose the same 35-year-old, whose waiting time is now one year; or would the choice be to switch to the five-year-old, whose waiting time is two years? As long as the same patient continued to be chosen for transplant now, the waiting time for this preferred patient was shortened by halves: six months, three months, and six weeks. This process is referred to as “halving”, and was continued up to four times. If the preferred patient continued to be chosen on the fourth halving, the preference was noted as an “absolute choice”.

There were four questions, where 35 was set as the reference age, and this was compared to ages 5, 10, 60, and 70. Then there were another four questions, where 60 was set as the reference age, and was compared to ages 5, 10, 35, and 70. Thus, all respondents were asked the same set of eight questions each.

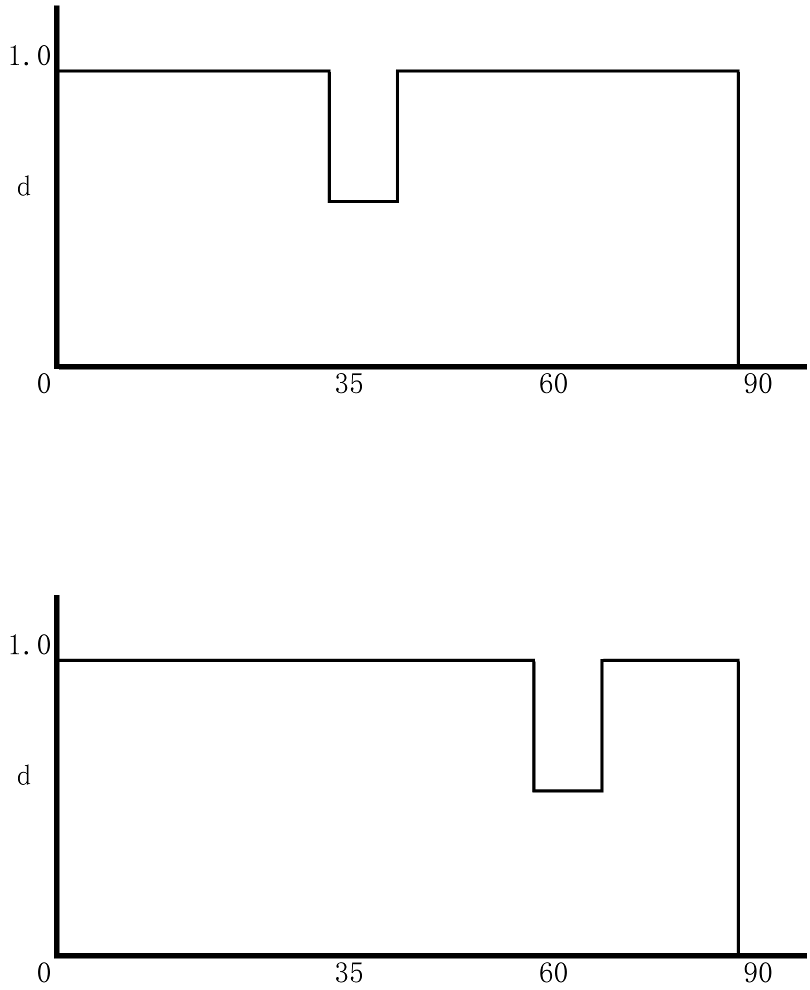
The main body of the interview script, as published in the BHC paper, was translated into Japanese by the author, with four minor changes.[3] The first change was that: while the BHC interviews only stated that the two patients would live to the same age, this age was specified to be 90. This was done to avoid the possible misunderstanding that the two patients will die at the same point in time, at different ages.

Secondly, visual aids such as the one shown in Figure 1 was used, which illustrates the HRQOL (Health Related Quality of Life) path each patient would follow, if he/she was not chosen for transplant now.[4] The horizontal axis represents patient age, and the vertical axis is for the level of

[3] This is available from the author on request.

[4] The figure is based on Figure 1 printed in Busschbach et al.(1993). Nevertheless, note that the original BHC study did not use these visual aides in the actual interviews (personal communication, Busschbach, October 1996).

Figure 1: The visual aid



health. *d* denotes the level of health at dialysis. (The figures were not labelled, and verbal explanations were given.) The top (bottom) panel of the figure corresponds to the health profile of the patient who is presently 35 years old (60 years old), if this patient was not treated now, and was to wait for two years on dialysis. There were additional figures with relevant combinations of patient age and waiting time that were used in accordance of the choices made by the respondent.

The third change was for those respondents who made absolute choices. While the BHC study simply dropped the absolute choices from further analysis as being invalid, in the replication study they were asked to confirm that their preference was indeed “absolute”. This was done by asking whether they will continue to choose the same patient even if the chosen patient had only one day to wait while the unselected patient had to wait for two years, if not treated today.

Further, after all the questions were over, the respondents were invited to talk freely on the decision criteria they used. (The time taken for this is not included in Table 2 below.)

2.2 Subjects and exclusion criterion

As in the BHC study, two groups of respondents, of different ages, were recruited. For the younger group, members of a students' club at Kyoto University were approached and asked to co-operate. As for the elderly group, two institutes agreed to ask their elderly members to be interviewed: one was an adult education institute and the other was a day-care service centre, both located in Kyoto. All the interviews were carried out by the author, and, as a reward, the respondents were given roughly $5- and $10-worth of "book tokens", depending on whether or not the respondent had to travel to the interview venue. One student and four elderly respondents declined to receive this.

Respondents who only gave absolute choices were excluded from the main body of analyses of the four hypotheses explained below. There were put back into the dataset for the additional analyses later on.

2.3 Data analysis

Following the BHC study, it was assumed in the replication study that the real switch point, or indifference point, can be approximated by the midpoint in the number of halvings. For example, if a respondent continued to prefer the 35-year-old up to two halvings and switched the preference to the five-year-old on the third halving, then this means that the respondent evaluates the value of a two-year-period starting from age five to be equivalent to less than six months, but more than three months, at age 35. Given 2.5 halvings = $2^{-2.5}$ = 0.18, assuming:

> that the indifferent point is at 2.5 halvings,
> a zero temporary discount rate, and
> that the change in the value of life during the two-year period from age five and the six-month period at age 35 is negligible relative to the difference between the two ages,

a year of life at age five is valued roughly 0.18 times that at age 35. It should be noted these assumptions are crude and the implied weights thus derived are of approximate nature.[5] If the respondent refuses to prioritise between the two patients (and for example suggests a lottery), then the point of indifference in terms of the number of halvings will be 0, and in terms of the implied weight will be 1 for either patient/age.

Note that this procedure involves a power transformation. Therefore it is impossible to assume that *both* the number of halvings *and* the implied preference value are based on symmetric distributions. The BHC study first calculated the average for switch points in terms of halvings, then converted these to relative preferences, and finally carried out the relevant statistical tests. The tests require the assumption that the underlying distributions of the relative preferences are symmetrical. However, the numbers of halvings observed in this study are more symmetric than their transformed implied weights, which have long right-hand tails. Therefore, the present study performed some of the statistical tests on averages in terms of number of halvings, and not of the implied weights.

2.4 Hypotheses and analyses

There were three hypotheses to be tested, which are explained in turn.

---

[5] A more elaborate approach will be to explicitly assume a continuous age-value function. For example, following Murray, Lopez (1994), suppose $v = x\ exp(-bx)\ exp(-r(x-a))$, where v is the value, x is age, a is the patient's present age, b is the age discount factor, and r is the temporal discount factor. The value of health for a fixed period, t, will be equal to the definite integral of this between $x = a$ and $x = a + t$. The example in the text can be translated as the case where the solution of this definite integral when $a = 5$ and $t = 2$ is equal to the solution when $a = 35$ and $t = 2 \times 0.18$. With temporal discount rates of 0%, 5% and 10%, the values of b that satisfy this relationship are 0.00123, 0.00272, and 0.00418 respectively. The implied relative weights with these parameters are 0.15, 0.16, and 0.16. The crude approach employed in this paper assumes no particular functional form and ignores temporal discounting.

2.4.1 The first hypothesis

The first hypothesis is:

H1: the relative value of health decreases with age.

Following the BHC study, this was tested in two ways. One was by fitting a straight line, using OLS, through the profile between patient age and the number of halvings under each reference age. If the relative value decrease with age, then this coefficient will be negative. The number of halvings and not the implied weight is used, since the distributions of the implied weights demonstrate a long right-hand tail. Secondly, under each reference age, the equivalence of the number of halvings is compared pair-wise, using a *t*-test: this will pick up hump-shaped profiles. Given the sparseness of the ages valued (5, 10, 35, 60 and 70), the main purpose is to determine the rough shape of the profile, and not to identify a continuous function that covers the entire life span.

Note that the decrease in life expectancy with age is irrelevant to the 1st hypothesis. The questions are designed so that the two candidate patients are not competing for length of life. The comparison is between a time slice at one age and another time slice at another age, and the issue is the relative value of improvement in HRQOL at these different ages.[6]

2.4.2 The second hypothesis

The second hypothesis is:

H2: this negative correlation (H1) is independent of respondent's age.

An ideal test of this would be to interview people of all ages and to regress their responses to their age, but such an approach would require a very large number of respondents, and therefore a different design was adopted. As is mentioned earlier, an elderly respondent group and a younger, student respondent group were asked to answer the same questions and their responses were compared, using *t*-tests.

2.4.3 The third hypothesis

The third hypothesis is:

H3: this age-related preference (H1) can be expressed on an interval scale.

For this hypothesis, the BHC study adopted the "parallelism method" (Forberg and Kane,1989). This will examine whether or not the preferences and the implied weights satisfy the interval scale properties so that there is scope to use the elicited age-related preferences as a basis on which to design age weights to apply to outcome measures of health in economic evaluation.[7]

The interview design where two reference ages (35 and 60) were used come into play at this stage. If the magnitude of age-related preferences for health was captured in interval scales, then the two profiles of implied weights with different reference ages should be vertically parallel to

---

[6] Put in a personal framework, the issue is parallel to considering the question: *supposing one had to go through a certain decline in HRQOL for a fixed amount of time, when would be the least/most harmful age for this to happen, and by how much?*

[7] If preferences, or weights derived from them, have ordinal scale properties, then the distance between values will have no meaning so that while it is correct to say values 0.2, 0.6, 0.8 are in ascending order, we cannot say that the difference between 0.2 and 0.6 is larger than the difference between 0.6 and 0.8. If the preferences and the weights derived are on an interval scale, not only is the former difference larger than the latter, the former difference is twice as large as the latter. Any set of weights should be expressed on a ratio scale, which is a special type of interval scales, so that satisfying the interval scale property is a necessary, but not sufficient condition for the design of age weights.

each other. For example, the implied weight for age five relative to age 35 and the weight for age 10 relative to age 60 are different because (*a*) the ages themselves (5 and 10) are different and (*b*) the reference ages (35 and 60) are different; and a crucial assumption is that there is no interaction between factors (*a*) and (*b*). If this is the case, then the weight profile obtained relative to age 35 and that relative to age 60 will be parallel to each other, so that the implied weight for age *x* relative to one age will be smaller or larger than that under the other reference age by a constant amount.

A so-called "multi-dimensional ANOVA (analysis of variance) in two-way layouts with repetitions" was carried out[8] to test for the extent of interaction between the two profiles, for each respondent group. This analysis breaks the observed difference within the data down to factors due to different age in question (factor *a*) and different reference ages (factor *b*), and looks for the possibility of interactions. Note that the null hypothesis of the ANOVA is that there is no interaction: ie. the smaller the *p*-value a better indication of the existence of interaction; however, a high *p*-value does not necessarily demonstrate that there is no interaction.

Ideally, an ANOVA will be carried out on data without skewness, which in this case is the number of halvings rather than the implied weights. However, since it is concerning the implied weights, and not the number of halvings, that we want to confirm the non-existence of interaction and thus the satisfaction of the interval property, the analysis is carried out on the implied weights. This will be performed for the two respondent groups separately.

Further, if the expressed preferences have a cardinal property, then the implied weight for age 60 with reference age 35 should be the inverse of the implied weight for age 35 with reference age 60.[9] Equivalently, the number of halvings for one of these questions will be equal to the reply to the other question, multiplied by –1, and *t*-tests are carried out to examine this.

2.5 Additional analyses

As is mentioned above, the BHC study excluded the absolute choices from all analyses. The reason for so doing was that, at four halvings, the ratio of the waiting times become two years : six weeks = $1 : 2^{-4} = 1 : 0.0625$, where it seemed more plausible to assume that the respondent has an absolute preference over the health of the chosen age/patient than to assume that a relevant trade-off was still being considered. There also was the practical reason that, in order to perform quantitative analyses based on average choices, absolute choices had to be omitted. Nevertheless, there were many absolute choices in this replication study, and an additional series of non-parametric analyses was carried out in order not to waste the ordinal preference information gathered. Since each respondent has two sets of data from the two reference ages, within-respondent consistency of the individual respondents was calculated. Further, their orderings were compared with the ordering obtained from the average preference of the relevant respondent group for within-group consistency.

3. RESULTS

3.1 Respondents and exclusions

Table 3 is a summary of the respondents. This differs from the original BHC study in two aspects. First, the number of total respondents was smaller: by 50% for the elderly group and by 30% for the student group.[10] Second, the average age of the elderly group was lower: the average age of the elderly group in the BHC study was close to 80 years old. This latter point means that there were more respondents who found themselves choosing between two hypothetical patients one of whom's age was about the same as his/her own. The extent of the possible effect of this element is unknown.

---

[8] MS Excel was used.

[9] I am grateful to the anonymous referee at CHE who suggested this test, which is not included in the BHC study.

[10] This was partly due to limitation in funding and time available, partly due to difficulty in recruiting elderly respondents.

Table 1: The respondents

	elderly	students
respondents total	24	21
retirees †	3	0
invalid throughout ‡	6	0
percentage	25	0
average age (yrs)	73.2	20.1
SD	6.9	1.8
time taken (min)	16.6	6.0
SD	6.3	3.0

† the number of respondents who did not complete the interview
‡ the number of respondents who only made absolute choices

The percentage of respondents who gave absolute responses throughout was the same as the BHC study, viz. 25% of the elderly group. These respondents were excluded from the analyses for the three hypotheses, but included for the later additional analyses.

3.2 Average responses and the hypotheses

Average responses are shown in Table 2. The following notation will be used hereafter: the respondent groups will be referred to with capital letters E (elderly) and S (students), and the reference age with /35 and /60: e.g. 10(S/60) means the average response of the student group for age 10 with reference age 60. Figures 2 and 3 illustrate the results by respondent group and reference age respectively. The three hypotheses are examined in turn below.

Table 2: Average responses

ref. age	age	elderly respondents				student respondents			
		relative age weight	average halving	SD †	valid re-sponses	relative age weight	average halving	SD †	valid re-sponses
35	5	0.610	-0.71	1.21	14	1.753	0.81	1.26	21
	10	0.807	-0.31	1.42	17	1.641	0.71	0.98	21
	60	0.525	-0.93	1.40	14	0.517	-0.95	1.38	21
	70	0.286	-1.81	1.86	13	0.552	-0.86	1.66	21
60	5	1.250	0.32	1.73	14	1.811	0.86	1.33	19
	10	1.492	0.58	1.39	13	2.282	1.19	1.44	21
	35	2.520	1.33	1.34	12	1.561	0.64	1.30	20
	70	1.065	0.09	1.06	11	0.992	-0.01	1.00	21

† SD is the standard deviation of the implied switch point in terms of the number of halvings

3.2.1 The relative value of health decreases with age (H1)

Table 2 shows that the response of the elderly group has a peak at age 35 under both reference ages, while the profile obtained from the student group is mostly downward sloping. Also see Figure 2. Table 3 shows the results of the linear regression with age as the independent variable and the implied average switch point as the dependent variable. While the student group's response clearly presents a negative profile, that of the elderly group is ambiguous.

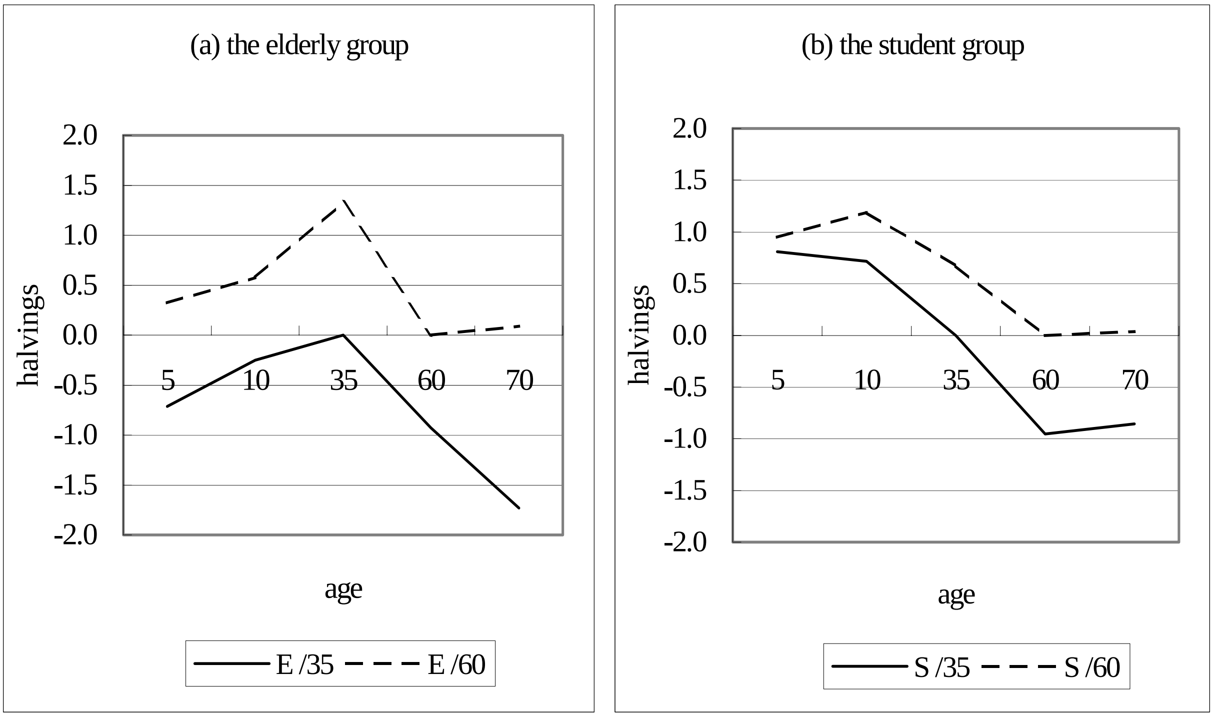


Figure 2: The results by respondent group

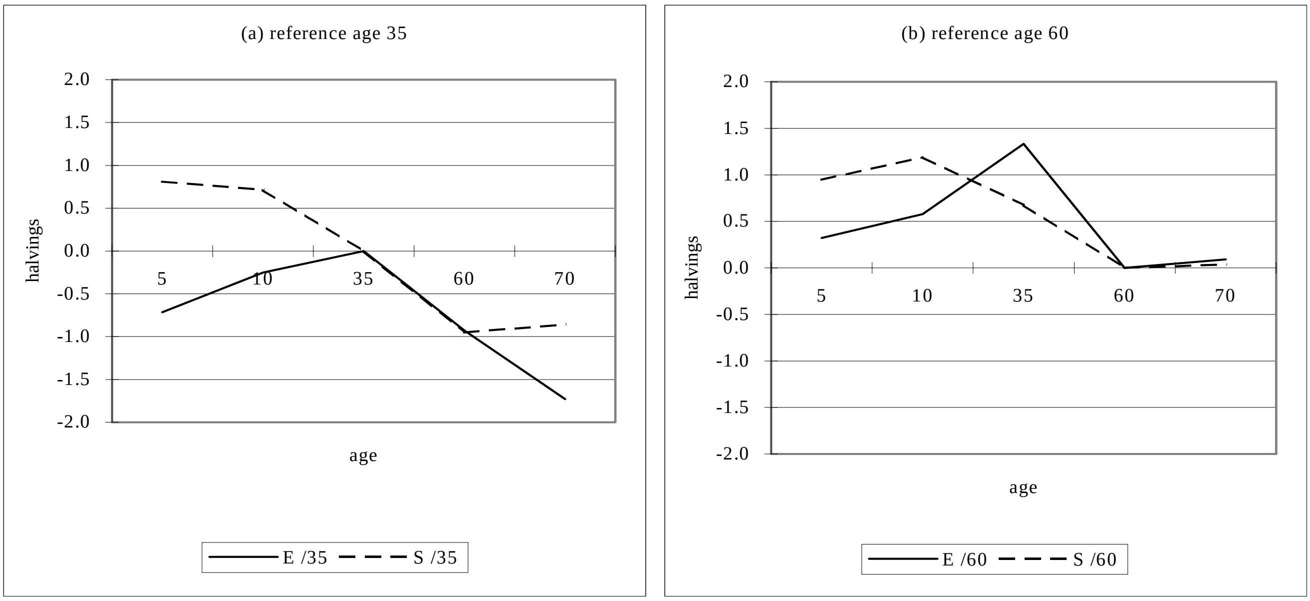


Figure 3: The results by reference age

Table 3: The results of linear regressions

		co-efficient	SE	*p*-value	r^2
the elderly /35	intercept gradient	-0.164 -0.016	0.439 0.009	0.733 0.213	0.439
/60	intercept gradient	0.712 -0.007	0.468 0.009	0.202 0.535	0.140
the students /35	intercept gradient	0.965 -0.028	0.121 0.003	0.004 0.002	0.973
/60	intercept gradient	1.208 -0.018	0.131 0.003	0.003 0.009	0.923

† independent variable: age
dependent variable: average switch point in terms of the number of halvings

Table 4 demonstrates for each of the respondent groups the result of *t*-tests, which compared different age pairs to pick up hump-shapes. A negative (positive) sign before the *p*-value here indicates that the slope between the age pair is negative. The upper-right hand half is for comparisons with reference age 35, and the lower-left hand half is for reference age 60. Thus, for instance, the value in the upper-right corner cell in panel (a) means that within the elderly group, and with reference age 35, the relative value for age 70 is smaller than that for age five, and that the relevant 1-sided *p*-value is 0.047.

Table 4: 1-tailed *P*-values from *t*-tests on the average switching point (H1)

(a) The elderly group

/35 /60	5	10	35	60	70
5		0.171	0.002	-0.329	-0.047
10	0.344		0.212	-0.112	-0.015
35	0.063	0.099		-0.002	<-0.001
60	-0.190	-0.026	<-0.001		-0.118
70	-0.356	-0.186	-0.015	0.343	

Upper half is for reference age 35 and lower half is for reference age 60.
The negative sign indicates that the slope between the two ages is negative.

(b) The student group

/35 /60	5	10	35	60	70
5		-0.396	-0.003	<-0.001	<-0.001
10	0.296		-0.001	<-0.001	<-0.001
35	-0.266	-0.124		-0.002	-0.013
60	-0.001	<-0.001	-0.013		-0.422
70	-0.010	-0.003	-0.046	0.437	

Upper half is for reference age 35 and lower half is for reference age 60.
The negative sign indicates that the slope between the two ages is negative.

It shows that for the elderly group, age 35 is clearly the peak of the profile amongst the five ages that were compared in the interviews, so that their age-value profile is hump-shaped. As for the student group, although *p*-values are not significant, an upward slope is observed between ages 5 and 10 and ages 60 and 70, both under reference age 60. Their age-value profile may be better described as an inverse *s*-shape, with a plateau up to 35 and then a decline that flattens out. Given the sparseness of the age of the patients, the exact profile cannot be determined beyond this crude observation.

3.2.2 The effect of respondent's age (H2)

As was seen in the section above, the responses of the two respondent groups generate different profiles. Figure 3 illustrates the average responses by reference age: 2(a) is for reference age 35, and 2(b) is for reference age 60. To test for the degree of the agreement, *t*-tests for overall average (under each reference age), and for each corresponding response pair were carried out.

The first *t*-test compared the average switch point for all ages by the elderly group under one reference age with that by the student group under the same reference age. This showed that the effect of respondent group on the results as a whole is not largely significant: p=0.089 for /35 and p=0.683 for /60 (2-tailed).

Nevertheless, the second *t*-test comparing the responses for each age indicates that the similarity is limited. As is illustrated in Table 6, while there is a strong agreement concerning the preference over the advanced ages, that for childhood on the other hand, especially under /35, is weak.

Table 5: *t*- tests on the effect of respondent groups (H2)

	/35				/60			
	5	10	60	70	5	10	35	70
p †	0.001	0.027	0.962	0.179	0.263	0.242	0.194	0.889

3.2.3 The age-related preference can be expressed in interval scales (H3)

The issue here was whether or not the profiles of implied weights are parallel to each other. The ANOVA for the extent of interaction between the weight profile under /35 and that under /60 resulted in *p*-values of 0.14 and 0.38 for the elderly and student respondent groups respectively. The implication is that the ANOVA failed to reject the null hypothesis that there is no interaction, especially for the responses of the student group. Further, when the number of halvings for 60/35 and 35/60 were compared, 46% and 60% of the elderly and student respondent groups satisfied the condition that one should be the negative of the other. The results of the *t*-tests for this are $p = 0.10$ and 0.27 for the elderly and student respondent groups respectively. Thus, there is little support for the interval property of the responses obtained from the elderly respondents, while there is more support for the responses obtained from the students.

3.3 Additional analyses including the absolute choices

The results obtained from the interviews have a relatively large degree of variance, suggesting that there may be additional insight gained by focusing on individual responses. In this section therefore, three issues concerning the individual responses are taken up.

3.3.1 Clear rejections of the 1st hypothesis

There was one respondent (student) who refused to prioritise between the two patients at the very first stage of the interview. The respondent's opinion was that, as a matter of medical ethics, provided that there were no differences in medical factors, attributes such as age, family status, and occupation should not count and the patient to be treated first should be chosen by a lottery.

Moreover, there were four respondents (three elderly, one student) who gave higher weights to older patients throughout (ie a complete rejection of H1). There were 28 responses (11 from the elderly, 17 from the students) which gave the 70-year-old patient a higher preference, compared to the 60-year-old patient.

3.3.2 Within-respondent consistency

It should be noted that many absolute choices were obtained in the interviews. While the BHC study interviewed 77 subjects and encountered 38 such responses, there were 66 such responses in this reproduction from 45 respondents, of which 64 came from the elderly respondent group.[11] When a respondent did not switch between patients after four halvings, the initial procedure, following the BHC study, was to register the response as an absolute choice and to omit it from hypothesis testing. Nevertheless, since these responses contain information regarding which of the two patients are preferred, an additional non-parametric analysis was carried out *including* these responses. Here, by calculating Spearman's rank order coefficient (*r*) between preferences under /35 and under /60, within-respondent consistency was calculated. In so doing, all absolute choices were assumed to be of the same ranking. Table 6 shows the distribution of respondents according to their rank order coefficient. The 3rd row of the table indicates that, those respondents, whose responses were largely absolute (viz., more than 6 absolute choices in 8 questions), do not necessarily coincide with those whose responses were inconsistent. Or, in other words, out of those respondents whose rank order coefficient is significantly low (r<0.7), 79% (11/14) have made less than 6 absolute choices and had their responses reflected in the analyses of the foregoing section.

Table 6: Distribution (in number of respondents) of within-respondent rank order coefficient (Spearman's *r*)

coefficient †	0 < 0.5	<0.7	<0.9	≦1.0	subtotal
elderly	1	8	5	7	21
of which, AC6 ‡	0	3	2	2	7
students	2	3	4	12	21

† 2-tailed 10% cut-off for *n*=6 is *r*=0.829
‡ AC6: Those with more than 6 "Absolute Choice" responses

3.3.3 Within-group consistency

Since a large portion of the responses in the elderly group were absolute choices and hence omitted from quantitative analysis *in*cluding the absolute choices may affect the overall rank ordering for this group. Indeed, it was found that the rank orders of many elderly respondents were not correlated with the rank order of the average response of their own group. Table 7 presents within-group consistency defined here as the rank order coefficient within a group, between the rank ordering of each respondent *in*cluding the absolute choices, and the rank

[11] Most of the absolute choices in the BHC study (32/38 = 84.2%) were in favour of the younger patient. From this, the BHC paper observes that, while these were excluded from further analysis, the direction of preference was in accord with the general trend of favouring the younger patient to the older. In the present study, three quarters of them (17/66 = 74%) were in favour of the younger patient.

ordering of the averages of the relevant group *ex*cluding these absolute choices. Note that each respondent has two sets of responses for each age (for the two reference ages) and hence the numbers in the column marked "subtotal" is double the number of respondents in each group. This table shows that 77% (33/43) of the elderly respondents cannot be said to share the ranking of the average of their group (r<0.9), and further, 14% (6/43) had negative rank orderings. On the other hand, 85% (36/42) of the students' responses are significantly positively correlated with the average of their group.

Table 7: Distribution (in number of respondents) of within-group rank order coefficient (Spearman's *r*)

coefficient †	＜0	＜0.5	＜0.7	＜0.9	≦1.0	subtotal
elderly	6	5	9	13	10	43
students	0	0	3	3	36	42

† 2-tailed 10% cut-off for *n*=6 is *r*=0.829

4. DISCUSSION

4.1 The nature of age-related preferences

What are the determinants of age-related preferences? The BHC paper provides three possible interpretations as to why people may value health differently depending on age:

First, it might be possible that health at a younger age has a higher utility because it has influences on the *development* of the patient. […] A second reason might be that at mid-life, one has more *responsibility for others* […]. A third reason might be that younger people are seen as more *valuable for society* […]. (Italics added.)

Note that these are based on the authors' understandings, and are not elicited from the respondents themselves.[12]

In the present study, after the interviews were over, respondents were invited to talk freely on what they thought most relevant and how they had reached their decisions. Most of the explanations that supported negative profiles were similar to those mentioned in the BHC paper. One prevailing exception was the argument that, those who have already lived longer should give way to those who still have not.[13] For an argument contrary to the negative slope profile, there was an argument that, as one grows older, the length of remaining life decreases, making each remaining year more precious, and hence older patients should be given priority over younger patients.

There seem to be several possible arguments, which may be supported singly or in combinations, regarding age-related preferences, and depending on them, the choice of patient to be treated first will change. The following five appear to be the main arguments:

[12] In the original BHC interviews, respondents were presented with several arguments before answering, which include the above three. Not included in the BHC paper were: "Being sick is worse for the young (or for the old)", "The old are in a more difficult position", and "The fair innings argument" (personal communication, Busschbach, October 1996). Note that (a) the first of these is incompatible with what the BHC study refers to as "The first assumption" of their survey: viz. "the ratio of the utility of a decrease of health is the same as the ratio for the utility of health itself", and (b).no arguments were presented to the respondents in the present study.

[13] There was not much of an "argument" given for this opinion, and those who supported it seemed to find it rather self-evident. Nevertheless this coincides with the "fair innings argument".

(1) give priority according to economic earning power and productive efficiency,
(2) give priority to the age where investment and learning is greatest,
(3) consider family and/or social responsibilities, or externalities,
(4) give priority to those who have lived less, or the "fair innings argument", and
(5) give priority to those who have less left.

4.2 Interpretation of absolute choices

As has been noted, the results include far more responses that were absolute than found in the original BHC study. Furthermore, while those in the BHC study may have included those who were to switch their preference on the next halving, all the respondents who declined to switch their preference on the 4th halving in the present study confirmed that their preference is such that they will give absolute priority to the patient selected first.

In order to quantify age-related preferences, it is obviously impossible to include preferences that are "absolute", and thus the BHC study excludes these responses from the analysis. They also point out that these responses "could be interpreted as representing the grey area between understanding and non-understanding" the interview task. Nevertheless, the additional non-parametric analysis shows that there is not much correlation between not making absolute choices and achieving consistent rank ordering (cf. Table 6). In other words, there were several respondents who have made absolute choices, but whose preferences were consistent. A possible explanation for the results of Table 7, where it was found that the rank ordering of 2/3 of the elderly respondents are not correlated with the ordering of the average of their group, is that the group average has been calculated *after* excluding their absolute-yet-consistent choices.

There are two conclusions to draw from this. The first is that there may be explanations other than poor understanding of the interview task for at least some of the absolute choices. The second is that, being largely inconsistent, rather than making absolute choices, is a more straightforward sign of poor understanding of the task, and accordingly, it seems inappropriate to produce average preferences by *in*cluding these inconsistent preferences, while *ex*cluding the absolute-but-consistent ones.

4.3 Absolute choice and the status-quo bias

There is a possibility that at least some of the absolute choices were affected by the so-called "status quo bias" (Samuelson and Zeckhauser,1988). Status quo bias refers to the observation that when people are given an opportunity to change a choice they have made earlier, they are likely to stick to the original choice well beyond the point where they would have chosen the alternative had they not already made the first choice. This implies that, for example, a respondent who, if asked directly to prioritise between a 10-year-old who has to wait for one year if not treated today and a 35-year-old who has to wait for two years if not treated today, would prefer to treat the 35-year-old first, may nevertheless continue to select the 10-year-old even when the 10-year-old needs only to wait for six months or less when the questions are asked in the halving procedure. This is because having selected the 10-year-old in the initial phase will then have become the status quo for the subsequent halving questions. The design of the interview, therefore, is such that there is scope to expect this bias, which will result in delaying the switch point and possibly increasing absolute choices. Thus, the quantitative results of both the BHC study and this replication are likely to be exaggerated, while the preference profile may well be more flat in the absence of the bias.

It seems natural to assume that this bias will be stronger when the decision context is on choosing between different persons than between different goods or portfolios (as was the case in the study reported in Samuelson and Zeckhauser,1988). It also seems natural to suppose that, while the students dealt with the tasks in an abstract manner, the elderly pictured more vivid images of the patients competing for treatment. It may thus be argued that the elderly are more vulnerable to the bias, and are therefore more likely to produce absolute choices.

This implies two methodological issues yet to be explored. First, HRQOL elicitation methods that involve interpersonal trade-offs may be more vulnerable to status quo bias than those methods that are confined to intra-personal trade-offs. Second, standard PTO, dealing with hundreds or thousands of people, may be less vulnerable than TTO employed here, trading off time between one person and another.

4.4 Conclusion and implications

Both the original BHC study and the reproduction study found that people value health differently according to the age of the patient. The two major limits of these studies are their small sample size and the possible existence of the status quo bias. Further, given the weak evidence regarding the interval property of the values generated in the present study, the results reported here should not be regarded as final in any sense. Nevertheless, that people evaluate a given unit of health differently according to the age of the patient, and that this age-value profile declines beyond middle age, seems to be a robust overall description.

More empirical research is needed to confirm this pattern, to specify the age-value profile in continuous form, and to narrow down on the implied age weights.

Further, theoretical research is required to explore (a) the ethical relevance of this age-related preference of the general public to actual health care resource allocation policies and, to the extent that it is ethically relevant, and (b) the methods by which this preference can be reflected in QALYs and in economic evaluation of health care interventions.

REFERENCES

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

Forberg, D.G., Kane, R.I. (1989), Methodology for measuring health-state preferences – I: Measurement strategies, *Journal of Clinical Epidemiology*, 41(2):345-354

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

Murray, C.J.L. (1996), Rethinking DALYs, in Murray, C.J.L., Lopez, A.D. eds., *The Global Burden of Disease*, Harvard University Press

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

Patrick, D.L., Bush, J.W., Chen, M.M. (1973), Methods for measuring levels of well-being for a health status index, *Health Services Research*, 8: 228-245

Samuelson, W., Zeckhauser, R. (1988), Status quo bias in decision making, *Journal of Risk and Uncertainty*, 1: 7-59

Tsuchiya, A. (1999), Age-related preferences and age weighting health benefits, *Social Science and Medicine*, 48(2):267-276

Tsuchiya, A. (2000), QALYs and ageism: Philosophical theories and age weighting, *Health Economics*, 9(1):57-68

Willliams, A.(1997), Rationing health are by age: the case for, *British Medical Journal*, 314:820-822

World Bank (1993), *World Development Report 1993: Investing in Health*, Oxford University Press