Conference or Workshop Item:

Tsuchiya, A. orcid.org/0000-0003-4245-5399 and Hukin, A. (2005) Dispersion vs. concentration of health benefits: preliminary report. In: Health Economists' Study Group (HESG 2005), 30 Jun 2005, Newcastle, UK.

Dispersion vs. concentration of health benefits: preliminary report

Amy Hukin and Aki Tsuchiya*
School of Health and Related Research, University of Sheffield

1. Introduction

If two health programmes both produced 100 person years, can we say their benefits are the same? Conventional economic evaluation of health care interventions assumes yes, they are equivalent. On the other hand there seems to be some intuition based on inequality aversion that people may wish to disperse health benefits rather than to concentrate it to a few, so that for instance an additional 5 years each to 20 people seems more preferable than an additional 50 years to just 2 individuals. However, it also seems absurd that for instance extending the life of a very large number of people, say about half a million, by one minute each should be regarded as equivalent to extending the life of one person by one year (365 days × 24 hours × 60 minutes = 525,600). In other words, dispersing health benefits beyond some limit, or a threshold, will probably diminish the value of the total benefit.

Patterns of dispersion and concentration have been found in empirical studies by Ubel et al (1996), Choudhry et al (1997), and Ubel et al (2000). The issue of specific utility functions and the level of thresholds were first explicitly explored in the heath field by Olsen (2000; also see Olsen, 1994), and subsequently Rodríguez-Míguez, Pinto-Prades (2002) proposed a method to actually identify the threshold level. The current paper is a quasi-replication of the study by Rodríguez-Míguez and Pinto-Prades (hereafter referred to as the RP study). In what follows, the RP study is briefly introduced, the changes made in the current study are explained, and then the results are reported.

2. Methods

2.1. Outline of the study design by Rodríguez-Míguez and Pinto-Prades

This section gives a very brief outline of the design used in the RP study; for further details, see the original publication. The RP study assumes that social welfare is a sum of individual utility from QALY gains, and where individual utility from marginal QALY gains are not constant. More specifically, the individual utility function is assumed to take the form

$$u(t_i) = \alpha_1 \exp(-\alpha_2 t_i) t_i^{\alpha_3}, \qquad [1]$$

* Corresponding author: e-mail <a.tsuchiya@sheffield.ac.uk>. This work is part of a project on "The relative societal value of health gains to different beneficiaries", funded by NICE and NCCRM. We are grateful to John Brazier, Paul Dolan, Jan Abel Olsen, and Anna Stevenson for their input, and to all the respondents for their participation. The usual disclaimer applies.

where u is utility and t_i ($i = 1,2, \dots n$) represents QALY gains to individual i, so that both positive and negative inequality aversion can be expressed. Positive inequality aversion is a preference for the dispersion of a fixed total benefit, and is associated with a negative value of $u''(t)$. Correspondingly, negative inequality aversion is a preference for the concentration, associated with a positive $u''(t)$. Thus, by exploring the value of t above which $u''(t)$ is negative and below which it is positive, the threshold value can be identified.

Imagine a reference programme that gives 10 additional years of life in full health to 10 people, denoted (10 years, 10). If inequality neutrality holds, then with zero discounting this programme and the following 100-person-year ***scenarios*** should produce the same level of individual utility and thus social welfare:

(1 year, 100),
(2 years, 50),
(5 years, 20),
(20 years, 5),
(50 years, 2).

However, if there is positive (negative) inequality aversion, then for the reference programme to be equivalent to each of these scenarios the number of people (p^*) each benefiting by 10 years has to be larger (smaller) than 10. This is in effect a person trade off exercise, and the RP study uses "choice bracketing" to operationalise this. The objective of the exercise is to identify the value of p^* that will make a respondent indifferent between a given scenario and (10 years, p^*). If inequality neutrality holds, then (with zero discounting) $p^* = 10$ should hold.

In order to operationalise the person trade off exercise and to identify the value of p^* for each scenario, the following series of 10-year ***reference programmes*** is set:

(10 years, 1),
(10 years, 3),
(10 years, 5),
(10 years, 8),
(10 years, 10),
(10 years, 12),
(10 years, 15),
(10 years, 18),
(10 years, 20).

As can be seen, these reference programmes do not generate the same sum of health benefits, and, provided the respondent supported the view that social welfare is increasing in the number of people treated other things the same, there is a dominant ordering so that the programmes further down the list are more preferable. In essence, the RP study aims to locate each of the five 100-person-year scenarios above along this ordered list. Would (5 years, 20) be judged to be better or worse than (10 years, 1)? Would it be located further up this ordered list and equal to say (10 years, 20), or further down and equal to (10 years, 3)? By bracketing and narrowing down the number of people in the reference programme, this will lead to the identification of the value of p^*.

Furthermore, if the value of p^* varies across the 100-person-year scenarios, then it will be possible to infer for each respondent a ranking of the scenarios by using the size of p^*. This is compared with a direct ranking of the scenarios, and the reference programme (10 years, 10) at the individual level and at the aggregate level.

An important assumption at this point is that $u(t) \times p = u(10) \times p^*$, where t and p each represent the number of years and persons in the scenario in question. By solving this for $u(t)$,

$$u(t) = 10 \times p^* / p, \qquad [2]$$

which allows the use of regression analysis to estimate the coefficients of equation [1], and in turn the identification of the threshold value.

2.2. Changes to the RP study design

There are seven main changes to the design used in the RP study. First of all, whereas the respondents of the RP study are students, the respondents of the current study are members of the general public.

Secondly, the RP study contacted the respondents over three sessions: session one to ensure participants understood the exercise, session two to carry out the main exercise, and session three to test for reliability over time. However, participants in the current study each attended one session only.

These two points lead us to think the “choice bracketing” method used in the RP study to identify indifference between scenarios may be too confusing. This consists of a series of choices between pairs of treatments laid out in table format, alongside instructions on what the respondent should do next depending on the answer for the choice in question (see Appendix of the RP study for a replication). Therefore, thirdly, a new mode of administration, based on a set of cards with different treatment pairs was used. Each 100-person-year *scenario* was treated as one *question*. Within each question, a set of nine cards was used. Each card had two treatments printed on them; “treatment A” represented the *scenario* in question and thus remained the same across all cards within a question; “treatment B” represented one of the nine 10-year *reference programmes*, taken from the ordered list explained above. For each question, respondents were asked to look at the set of cards and place them in three different batches: one where treatment A is better than treatment B, one where two treatments are equally good, and one where treatment B is better than treatment A. An advantage of this mode of administration is that there is ample opportunity to test for a respondent’s consistency within and across questions.

The fourth change concerns the scenarios. The RP study used five 100-person-year scenarios presented above, with the number of years ranging from 1 to 50. However, it was felt that shorter durations needed exploring as well, and thus two scenarios (9 months, 150) and (6 months, 200) was added, and the scenario (2 years, 50) was dropped. See the next section for further details about the questions asked at different stages of the study.

Fifthly, the RP study assumed that all patients were 20 year-olds. However, in order to explore whether the preference for dispersion and concentration is independent of patient age, in addition to the six main questions assuming all patients are 20, two further questions were introduced, one where patient age was set at 10, and another at 60. The 100-peron-year scenario (5 years, 20) was used in both questions.

Person trade off is a method that uses the number of people as the vehicle of trade off, and it may be affected by the level of this nummeraire. More specifically, when a respondent prefers a treatment involving "5 patients" does this figure have an absolute value, regardless of whether it is 5 patients out of a pool of 200, or a pool of 2000; in other words would the reference group have an effect? Furthermore, would the same patient who chose 5 patients out of 200 also choose 50 patients out of 2000; in other words, would constant proportional *person* trade off hold? The sixth change explored these issues. All questions explicitly stated that all those treated were drawn from a pool of 200 patients (and the number of those treated in the scenarios ranged from 2 to 200), except for two additional questions. One of them stated that the pool of patients was 2000, but all other figures remained the same: this will test for the reference group effect. The other scaled up all the numbers of people (ie. the pool of patients and the number of those treated; and thus the person-years) by 10: this will test for constant proportional person trade off.

One concern over the RP design was that it used years of life in full health, and not QALYs. 100 person-years in full health is 100 QALYs. And not only can 100 QALYs be made up of different combinations of years of life in full health (eg. 5 years to 20 people; 10 years to 10 people; etc), if the QALY concept holds in this context, each of them should be equivalent to different combinations of number of years, health related quality of life, and numbers of people (eg. 5 years in full health to 20 people; 5 years in 50% QOL to 40 people; 10 years in 50% QOL to 20 people; etc). However, it was considered to be too complicated to explore this issue in detail. Instead, the seventh change consisted of a simple test of whether or not the preference for dispersion or concentration was affected when the outcome of the two treatments were less than full health (but the same across the two treatments). One question was added, where the scenario in question was specified as (40 years, 50% QOL, 5), which is comparable to (20 years, 5). Furthermore, three questions were introduced which are: (20 years, "slight mobility problems", 5), (20 years, "slight pain", 5), and (20 years, "mild depression", 5).

2.3. Method of data collection

The survey was carried out in group sessions of about 10 participants each held in University seminar rooms. The sessions began with some general descriptive information about the nature of scarce health care resources and the need for prioritisation at a macro level; and how current policy is based on concerns for maximising person years with no particular concern for the distribution of it across identical patients.

Participants were then broken up into smaller groups of 2 to 5, each lead by a member of the research team. This facilitated closer monitoring of how the participants progressed with the exercise. The participants were given a questionnaire booklet consisting of 16 questions across 16 pages, accompanied by corresponding sets of 9 cards as presented above, and an explanation of how the questionnaire worked. Table 1 summarises the questionnaire. To avoid confusion, the cards and the pages of the booklet were colour-coded so that for instance all the green cards were used for the question printed on the green page, and so on. Each page in the booklet had 3 boxes printed: 'I prefer treatment A', 'Treatments A and B are equally good' and 'I prefer treatment B', so that participants can use these to group the nine cards as explained above. All cards were placed in a random order and participants were asked to shuffle the cards each time they were used. Once participants completed this for all 9 cards they were then advised that they could swap the cards into different boxes as they saw fit and then asked to write down which cards they had place in each box onto the booklet.

Once all participants completed this for the first question, the whole group moved to the next question, and the same steps were repeated for the following 11 questions. After question 11 the group had a brief open discussion about the questionnaire and the topic. Once all smaller groups reached this point, a brief explanation was given to the wider group on the notion of health related quality of life, and how living for a longer number of years and poorer health might be equivalent to living for a shorter duration in better health. Participants then broke up into the smaller groups and followed the above steps for remaining questions 12, 13, 14 and 15.

After these person trade off exercises, question 16 was a direct ranking exercise, where participants were provided with 7 cards each representing the 100-person-year scenarios used in the 6 person trade off exercises, and the reference programme (10 years, 10). Respondents were asked to rank them form 1 (most preferred) to 7 (least preferred). The aim of this exercise was to analyse the extent to which the implied ranking obtained from the person trade off will coincide with the results of the direct ranking.

Since, to the best of our knowledge, this was the first time this card method had been used to measure the indifference value of concentrating or dispersing benefits the design was adapted throughout the sessions. As the length of each session was unknown during the first 2 sessions (N=20) participants were given only 9 questions: (1 year, 100), (2 years, 50), (5 years, 20), (20 years, 5), (50 years, 1), (5 years, 20), the age 60 question, the age 10 question, (5 yeas, 200) out of a pool of 200, (5 yeas, 200) out of a pool of 2000. In other words participants answered questions 3 – 8.

In the third session (N=12) 5 questions were added: (0.5 years, 200), (40 years, 50% QOL, 5) and 3 questions on scenarios with specific health problems; and instead (2 years, 50) (i.e. question 4) was removed. A further direct rank ordering exercise (question 16) was added. The 4 cards ranked included: (5 years, 20), (20 years, 5), (50 years, 2), (10 years, 10).

In the final 4 sessions (N=26) an extra question on (0.75 years, 150) was also included to gage more fully people's preferences over smaller *t*'s. Question 16 expanded to include all the scenarios.

There were opportunities throughout the meetings for participants to voice any opinions that they might have had regarding the task. Sessions lasted between 60 and 90 minutes, depending on the number of questions, the speed of the respondents and amount of queries and opinions voiced throughout the meeting. Sessions were audio-taped with the participants consent and brief notes were taken throughout to aid clarity when interpreting the results. At the end, participants were asked to fill in a background questionnaire, and to write down their thoughts regarding the content of the group sessions. Participants were given a payment of £15 for participating in the study and thanked.

2.4. Data analysis

2.4.1. Inconsistencies and exclusions

The analysis involves attaching for each 100-person-year scenario by respondent a score, or a position relative to the ordered list of 10-year reference programmes, and observations have to be dropped if these cannot be assigned unambiguously due to inconsistency. There are three kinds of inconsistencies. First, "*across* question inconsistency" arises, if a respondent places all cards in the box where treatment A (or B) is preferred throughout the main questions: it is not inconsistent to say that for example scenario (2 years, 50) is better than any of the reference programmes (eg. because only the number people matter, and not person years), but then this will be inconsistent with saying that scenario (50 years, 2) is also better than any of the reference programmes. In such cases all observations from the same respondent is removed.

Second, "*within* question inconsistency" arises, if a response indicates that a given scenario was better than say (10 years, 12) but worse than (10 years, 8), and this observation has to be excluded. However, when the same apparently inconsistent response was given in the less than full heath questions, these were not excluded, since they are in line with the maximal endurable time hypothesis (Sutherland et al, 1982; Stalmeier et al, 1996), and not necessarily illogical.

Third, there were several questions where another type of within question consistency could be tested, if participants choose treatments that had less health benefits rather than treatments that had more health benefits to the same number of people; for example Treatment A is preferred when Treatment A is (5 years, 20) and Treatment B is (10 years, 20). Individual responses with this type of inconsistency will be excluded.

2.4.2. The implied ranking from person trade off and direct ranking

For each individual respondent, an implied ranking of the 100-person-year scenarios were obtained from the results of the main person trade off exercises, and this was compared against their own direct ranking of the scenarios using Spearman rank order correlations. The mean coefficient across respondents is reported. Note that some respondents (N=12) give ranking data for 5 scenarios and others (N=26) for 7.

For each scenario, the average implied rank score and the average direct rank score were calculated across individual respondents, and the Spearman and the Kendal correlation coefficients are reported.

2.4.3. The results of the main questions

The value of p^* was identified for each observation. When participants indicated that treatment A was better with respect to some 10-year reference programmes and treatment B was better with respect t other references, but did not directly reveal an indifference value, or if they demonstrated that they were indifferent on more than one programme, the intermediate value of the interval was used. If participants always chose pairs with the greatest number of patients or the pairs with the greater number of years (i.e. thought treatment A or alternatively treatment B was better in all instances) an indifference value was inferred (i.e. in questions 1-9 and 12-15 if treatment A was always preferred $p^* = 20$ and if treatment B was always preferred $p^* = 1$; in questions 10 and 11 if treatment A was always preferred $p^* = 200$ and if treatment B was always preferred $p^* = 10$). This inferred indifference value was used when participants did not make trade-offs.

As was explained above, the crucial value of p^* with zero discounting is 10, and thus a series of 2-sided *t*-tests are carried out for each question. If mean $p^* = 10$, then the implication is that inequality neutrality holds and that on average people are indifferent between the scenario in question and the reference programme (10 years, 10).

For each pair of scenarios, whether the average preference is to disperse 100-person-years or to concentrate is reported. For a given pair of scenarios, if the scenario with the larger number of persons is associated with a larger mean p^*, then this is regarded as a preference for dispersion. See for instance the pair (1 year, 100), where mean p^* is 15.49, and (5 years, 20), where mean p^* is 12.85. The former scenario has a larger number of persons (100 > 20) and a larger mean p^* (15.49 > 12.85), and therefore the cell representing this pair of scenario is marked with an "S" for spreading, or dispersion. The opposite relationship holds for concentrating preferences.

Finally, values of $u(t)$ at the individual level are calculated using equation [2] and based on individual values of p^*. This is used to estimate equation [1], and subsequently to identify the level of t where $u''(t) = 0$, which is the dispersion-concentration threshold.

2.4.4. The results of the additional questions

The effect of patient age, the effect of reference groups and constant proportional person trade off, and the effect of less than full health are explored in terms of mean p^* and *t*-tests. Note that while these can establish whether or not the preference between concentration and dispersion observed at the main questions also apply at these different settings, we cannot draw conclusions about the robustness of the dispersion-concentration threshold value from these results.

3. Results

<u>3.1. Recruitment and background characteristics of respondents</u>

The study was conducted on members of the general public recruited in the city centre of Sheffield. Out of the 96 individuals that agreed to attend the sessions 68 participants (71%) actually took part across 7 sessions. The background details of the respondents are shown in table 2.

The background questionnaire also listed 5 negative and 5 positive adjectives that may describe the experiences and thoughts that the respondents might have of the session, and invited respondents to choose as many as they thought applied. Results showed that the majority of participants thought that the sessions were: "interesting"

(*N* = 57); “insightful” (*N* = 32); “informative” (*N* = 20); “enjoyable” (*N* = 21); and “too sensitive a topic” (*N* = 7).

3.2. Inconsistencies and exclusions

Of the 68 participants, 49% had no inconsistencies, and a further 24% only one inconsistency throughout the exercise. However, 2 participants (2.94%) were inconsistent across questions and were excluded from the analysis, and a further 8 participants (11.8%) violated the first type of within question in consistency on more than 8 questions, and were also dropped. In terms of specific observations, 6 responses to questions (5 years, 20), (5 years, 20, age 10) and (5 years, 200, out of 2000) were excluded due to the second type of within question inconsistency. (See table 3 for the distribution of the violation; figures also include respondents excluded for the first two inconsistencies, and thus add up to more than 6.) The same kind of results was observed for less than full health scenarios (see Table 4), but these were not excluded, since they are not necessarily illogical.

3.3. Rank order correlation between implied ranking from PTO and direct ranking

When individual level correlation between the implied ranking and the direct ranking is compared, mean Spearman rank order coefficients when the number of scenarios is 5 and when this is 7 are 0.080 and 0.47 respectively. The Spearman and the Kendal correlation coefficients for the average implied rank and the average direct rank range from 0.81 to 0.99 (see Tables 5 and 6, for different numbers of scenarios ranked). Results are comparable to those reported in the RP study (KCC = 0.86; SCC = 0.94; individual SCC (mean) = 0.81).

3.4. Results of the main questions

3.4.1. The value of p^*
Table 7 summarises the mean value of p^* derived from the main questions for the 100-person-year scenarios, and the *t*-test results. They suggest that the inequality neutrality assumption is rejected, so that on average people are concerned about how a fixed benefit of 100 person years is distributed across the pool of 200 patients. The results also illustrate that the value of mean p^* starts below 10 persons for (6 months, 200) increases to around 15 persons at (1 year, 100), and then declines to around 5 persons, with zero discount. The pattern for durations over 1 year is consistent with positive time preference, but not the pattern under 1 year. Figure 1 illustrates this, alongside the results reported in the RP paper: the height of the bar at 10 years is fixed at 10 persons for both studies, since we assume that (10 years, 10) is equivalent to (10 years, 10). Both studies show the same hump shaped pattern, but note that the two studies cover different ranges of duration.

3.4.2. Dispersion or concentration?
Table 8 summarises combinations of scenarios, and indicate whether respondents prefer dispersion or concentration of benefits. This shows that, for example, if the choice is between giving 5 years to a larger smaller (20) or giving 2 years to a larger number (50) then the average preference to disperse rather than to concentrate; or between giving 6 months to a larger number (200) or to give 5 years to a smaller number (20) then the average preferences is to concentrate rather than to disperse.

The concentration of *C*s in the upper left hand corner and the *S*s in the lower right hand corner is consistent with the bar chart shown in Figure 1, and is similar to the results of the RP study.

3.4.3. The regression results and the threshold value
Equation [1] is estimated as a linear function (see Box 1), resulting in:

$$\hat{u}(t) = 0.941\exp(-0.029t)t^{1.173}.$$

The value of *t* at which the second derivative of this equation is zero represents the threshold value between dispersion and concentration of benefits, and corresponds to *t* = 2.6 years (assuming no temporal discounting; with positive discounting this value will be higher). This value is considerably smaller than the comparable value reported in the RP study, which is 9.1 years.

3.5. The results from the additional questions

Table 9 shows that the values of p^* are statistically significantly larger than 10 across the three ages: in other words, if the choice is between giving 5 years to a larger number (20) or giving 10 years to a smaller number (10) then the preference to disperse rather than to concentrate is robust. There also seems to be a pattern across the three ages so that this preference is stronger when patients are older, but the results for 10-year olds and 60-year olds are not statistically significantly different from the results for 20-year olds.

As can be seen in the second and third rows of Table 10, values of p^* are statistically significantly larger than 100 across the two additional questions: in other words, if the choice is between giving 5 years to a larger number (200) or giving 10 years to a smaller number (100) then the preference to disperse rather than to concentrate is robust across different patient numbers overall. The comparison between the second and third rows represents the effect of reference group: ie. does it matter whether the 200 patients are from a pool of 200 or from a pool of 2000? The results cannot reject that it does not matter. The comparison between the first and the third rows represents constant proportional person trade off: ie. if all relevant numbers are scaled up by ten, would the value of p^* also be scaled up by ten? The results cannot reject that constant proportional person trade off holds.

Table 11 illustrates that when the total health benefit of 20 years in full health to 5 persons is substituted with 40 years in 50% health to 5 persons, p^* remains statistically significantly smaller than 10: ie. to disperse across 10 persons than to concentrate on 5. The last column indicates that the two values of p^* are not significantly different from each other, so it cannot be rejected that the preference for dispersion of health benefits in year of life in full health can be generalised to health benefits in QALYs, provided the composition of QALYs is equal across the two alternatives (ie. in this case, all treatments in this question had 40 years in 50% QOL).

As can be seen in the three bottom rows of Table 12, the values of p^* are statistically significantly smaller than 10 across slight mobility problems, slight pain, and mild depression; in other words, if the choice is between giving 20 years with a given health problem to a smaller number (5) or giving 10 years with the same health problem to a larger number (10) then the preference to disperse rather than to concentrate is robust across these different health problems. The last column indicates that the values of p^* from the three questions is not different from that obtained in the full health case.

4. Conclusion and discussion

Current decision making criteria does not take into consideration how a given fixed total health benefit is distributed across a group of people. If several treatments all generate 100 person years (or 100 QALYs), then their benefits are assumed to be equivalent. However, this practice may not mirror the preferences held by members of the public. It has been pointed out that (i) people may prefer to disperse the fixed benefit across a larger number of patients rather than to concentrate it to a smaller number of patients, but that (ii) this may be subject to a threshold so that the size of health gain to an individual patient is not too small. This paper is a preliminary report of a study that explored this dispersion-concentration relationship. The theoretical approach of an existing study carried out in Spain with student respondents (the RP study) was adapted, with several changes to the study design.

The main finings of this study are: firstly, that the newly developed card-based person trade off exercise to assess the 100-person-year scenarios against a set ordered list of reference programmes is feasible with a reasonable level of consistent responses. While those taking part found the topic challenging, they did not have complaints or suggestions about the actual mode of administration. We are planning to look into the background characteristics of those who committed more inconsistencies and see if there are any patterns.

Secondly, different values of p^* were observed for different 100-person-year scenarios, indicating that the assumption of distribution neutrality does not hold. On average, respondents preferred to disperse the 100 person years but not when the size (duration) per patient was small; this threshold was identified at 2.6 years (with zero discounting). We are planning to recalculate the threshold with different discount rates. The issue of framing is also worth exploring: ie. whether the threshold level is a function of the set of scenarios used (see Figure 1).

Thirdly, the results of the additional questions were favourable towards the general assumptions of economic evaluation: none of the following changes resulted in statistically significantly different values of p^*: changing the age of patients from 20 years old to 10 years old or to 60 years old; changing the size of the reference group from 200 to 2000; scaling up the whole scenario by ten; and using various kinds of less than full health. This study can be used as a pilot to calculate the necessary sample size to explore these issues further.

References

Choudhry, N, Slaughter, P, Sykora, K, Naylor, D. (1997) Distributional dilemmas in health policy : large benefits for a few or smaller benefits for many? *Journal of Health Services Research and Policy*. 2, 212-216.

Olsen JA. Persons Vs Years - 2 Ways of Eliciting Implicit Weights. (1994) *Health Economics* 3 (1):39-46.

Olsen, JA (2000), A note on eliciting distributive preferences for health, *Health Economics*, 19: 541-550

Rodíguez-Míguez E, Pinto-Prades JL (2002) Measuring the social importance of concentration or dispersion of individual health benefits, *Health Economics*, 11: 43-53

Stalmeier PFM, Bezembinder TGG, Unic IJ. (1996) Proportional heuristics in time tradeoff and conjoint measurement. *Medical Decision Making*. 16(1):36-44.

Sutherland HJ, Llewellyn-Thomas H, Boyd NF, Till JE. (1982) Attitude toward quality of survival : the concept of maximal endurable time. *Medical Decision Making*. 2:299-309

Ubel PA, Dekay ML, Baron J, Asch DA. (1996) Cost-effectiveness analysis in a setting of budget constraints - Is it equitable? *New England Journal of Medicine* 334(18):1174-7.

Ubel PA, Baron J, Nash B, Asch DA. (2000) Are preferences for equity over efficiency in health care allocation "all or nothing"? *Medical Care* 38(4):366-73.

Tables

Table 1. Questions asked to participants and different variants used.

Different variants	Treatment A (Health gain, patients)
Baseline variant:	
Question 1	(6 months (0.5years),200)
Question 2	(9 months (0.75years),200)
Question 3	(1 year, 100)
Question 4	(2 years, 50)
Question 5	(5 years, 20)
Question 6	(20 years,:5)
Question 7	(50 year, 2)
Age variant:	
Question 8. All patients aged 60	(5 years, 20)
Question 9. All patients aged 10	(5 years, 20)
Pool size variant:	
Question 10. Pool size = 200	(5 years, 200)
Question 11. Pool size = 2000	(5 years,200)
Less than full health variant:	
Question 12. Health gain means living in a 0.5 QOL	(20 years,5)
Question 13. Health gain means living with slight mobility problems	(20 years, 5)
Question 14. Health gain means living with slight pain	(20 years, 5)
Question 15. Health gain means living with mild depression	(20 years, 5)

Table 2. Background characteristics of respondents

Background questions	Number of participants
Sex: Male Female	28 39 (Mean = 35.3, SD = 0.50)
Employment Status: Full-time work Part-time work Student Seeking work Home-maker Retired	10 14 30 2 2 9
Highest qualification: GCSE/O level NVQ A level Degree Postgraduate qualification No formal qualifications	6 5 20 15 11 9
Has previously or presently worked for NHS	12
Has previously or presently been a main carer for a disabled family member	12
Children: 0 children 1 child 2 children 3 children	51 5 4 1
Perceptions of own QOL:	Ranged between 40% and 100% (Mean = 85.42, Median = 90.00, SD = 13.89)

Table 3. Frequency of respondents that chose treatments that had less health benefits rather than more health benefits to the same number of people.

The treatment pairs (Health gain, patients) vs (Health gain, patients)	Number of participants that chose lesser health benefits
(5 years, 20) vs. (10 years, 20)	3
(5 years, 20) vs. (10 years, 20); age 60	4
(5 years, 20) vs. (10 years, 20); age 10	5
(5 years, 200) vs. (10 years, 200)	3
(5 years, 200) vs. (10 years, 200); out of 2000	2
(20 years, 5) vs. (10 years, 5)	2
Total	19

Table 4. Frequency of respondents that chose treatments that had less life years in less than full health states than treatments that had more life years in less than full health over the same amount of people (maximal endurable time)

The treatment pairs (Health gain, patients) vs (Health gain, patients)	Number of participants that chose lesser health benefits
(40 years, 5) vs. (20 years, 5); living in 50% QOL	5
(40 years, 5) vs. (20 years, 5); living with mobility problems	1
(40 years, 5) vs. (20 years, 5); living with slight pain	2
(40 years, 5) vs. (20 years, 5); living with mild depression	3
Total	11

Table 5. Ranking of health programmes (session 3) [a]

S-PTO (Health gain (years), patients)	S-DO (Health gain (years), patients)
1, 100	1, 100
2, 50	2, 50
10, 10	10, 10
5, 20	5, 20
50, 2	50, 2

[a] From more to less preferred (mean). *n*=12; KCC = 0.99; SCC = 0.99; individual SCC (mean) = 0.80.

Table 6. Ranking of health programmes (sessions 4-7) [a]

S-PTO (Health gain (years), patients)	S-DO (Health gain (years), patients)
1, 100	1, 100
5, 20	0.75, 150
0.75, 150	5, 20
10, 10	0.5, 200
0.5, 200	10, 10
20, 5	20, 5
50, 2	50, 2

[a] From more to less preferred (mean). *n*=26; KCC = 0.81; SCC = 0.93; individual SCC (mean) = 0.47.

Table 7. Assessment of health programmes

Health gain (years), *t*	Number of patients, *p*	*Mean p**[a] (2-sided *t-test*)[b]
0.5	200	9.07 (0.25)
0.75	150	11.32 (0.310)
1	100	15.49 (0.001)
2	50	14.83 (0.001)
5	20	12.85 (0.001)
10	10	10
20	5	8.32 (0.001)
50	2	5.52 (0.001)

[a] Number of patients who would have to receive a 10 life-year increase in order that this programme be indifferent to the (*t,p*) programme.
[b] H_o: mean $p^* = 10$; mean p^* /=10; $n = 58$.

Table 8. Distributive preferences based on individual gain (years)[a]

(Gain in years, patients)	0.75, 150	1, 100	2, 50	5, 20	10, 10	20, 5	50, 2
0.5, 200	*C*	*C****	*C****	*C****	*C*	*S*	*S****
0.75, 150		*C****	*C***	*C*	*C*	*S****	*S****
1, 100			*S*	*S****	*S****	*S****	*S****
2, 50				*S**	*S****	*S****	*S****
5, 20					*S****	*S****	*S****
10, 10						*S****	*S****
20, 5							*S****

[a] *C:* preferences for concentrating; *S:* preferences for spreading;
***significance at 1% level; ** significance at 5% level; *significance at 10% level; *n* = 58.

Table 9. The effect of patient age

Gain in years, *t* (age of patients)	Number of patients, *p*	*Mean p**[a] (2-sided *t-test*)[b]	Effect of age (2 sided *t*-test)[d]
5 (age 60)	20	13.84 (0.001)	(0.49)
5 (age 20)[c]	20	12.85 (0.001)	-
5 (age 10)	20	12.55 (0.001)	(0.82)

[a] Number of patients who would have to receive a 10 life-year increase in order that this programme be indifferent to the (t,p) programme (n = 58).
[b] H_o: mean p^* = *10;* mean p^* /=10*;* n = 26.
[c] This row represents the corresponding row in Table 7 above.
[d] H_o: mean p^* = mean p^* for age 20

Table 10. The effect of different frames

Health gain (years), *t*	Number of patients, *p*	*Mean p**[a] (2-sided *t-test*)	Effect of frame (2 sided *t*-test)
5[c]	20 out of 200	12.85 (0.001)[b]	-
5	200 out of 200	124.15 (0.034)[d]	(0.579)[e]
5	200 out of 2000	131.72 (0.007)[d]	(0.756)[f]

[a] Number of patients who would have to receive a 10 life-year increase in order that this programme be indifferent to the (t,p) programme.
[b] H_o: mean p^* = *10;* mean p^* /=10*;* n = 26.
[c] This row repeats the corresponding row in Table 7 above.
[d] H_o: mean p^* = mean $p^* \times 10$ for 20 out of 200
[e] H_o: mean p^* = mean p^* for 200 out of 2000
[f] H_o: mean p^* = mean p^* for 20 out of 200

Table 11. The effect of less than full health

Health gain (years), *t*	Number of patients, *p*	*Mean p** [a] (2-sided *t-test*)	Effect of QOL adjustment. (2 sided *t*-test)
20 [c]	5	8.32 (0.001)	-
40 (1/2 QOL)	5	7.33 (0.001)	(0.10) [d]

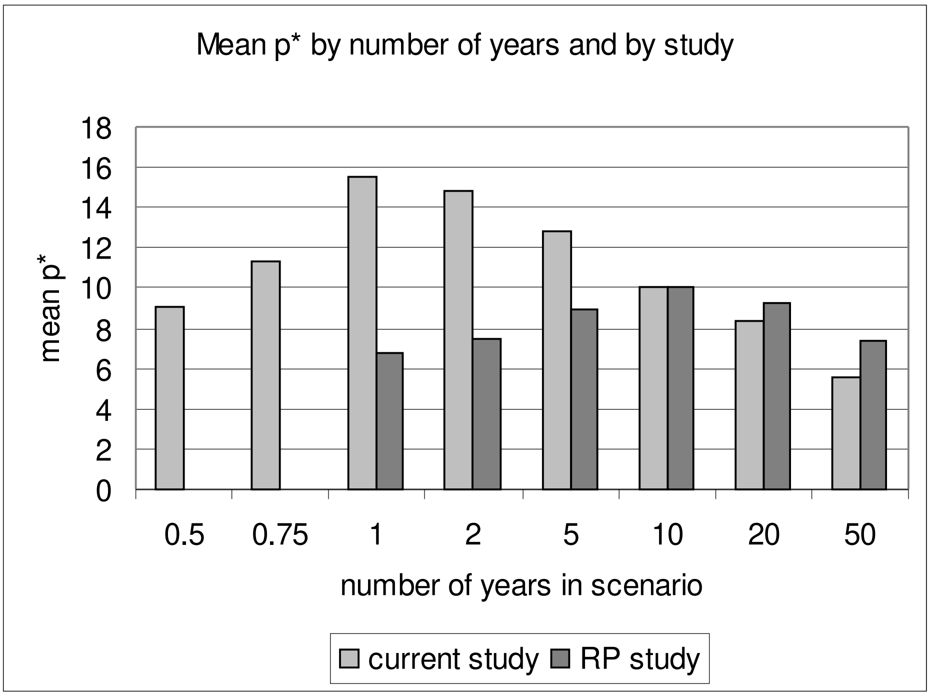
[a] Number of patients who would have to receive a 10 life-year increase in order that this programme be indifferent to the (*t,p*) programme.
[b] H_o: mean *p** = *10;* mean *p** /=10*;* *n* = 26.
[c] This row repeats the corresponding row in Table 7 above.
[d] H_o: mean *p** = *p** for (20 years, full health, 5)

Table 12. The effect of different types of ill health

Health gain (years), *t*	Number of patients, *p*	*Mean p** [a] (2-sided *t-test*)	Different health problems. (2 sided *t*-test)
20 [c]	5	8.32 (0.001)	-
20 (slight mobility problems)	5	8.47 (0.001)	(0.81) [d]
20 (slight pain)	5	7.69 (0.001)	(0.14) [d]
20 (mild depression)	5	7.84 (0.001)	(0.40) [d]

[a] Number of patients who would have to receive a 10 life-year increase in order that this programme be indifferent to the (*t,p*) programme.
[b] H_o: mean *p** = *10;* mean *p** /=10*;* *n* = 26.
[c] This row repeats the corresponding row in Table 7 above.
[d] H_o: mean *p** = *p** for (20 years, full health, 5)

Figures



Box 1: $\ln u(t) = \alpha_1 + \alpha_2 t + \alpha_3 \ln t + \varepsilon$

```
      Source |       SS       df       MS              Number of obs =     285
-------------+------------------------------           F(  2,   282) = 1561.99
       Model |  500.368547     2  250.184274           Prob > F      =  0.0000
    Residual |  45.1680106   282   .16017025           R-squared     =  0.9172
-------------+------------------------------           Adj R-squared =  0.9166
       Total |  545.536558   284  1.92090337           Root MSE      =  .40021

------------------------------------------------------------------------------
        lnut |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           t |  -.0270383   .0025426   -10.63   0.000    -.0320433   -.0220333
         lnt |   1.137616   .0292212    38.93   0.000     1.080097    1.195135
       _cons |  -.0135368   .0395854    -0.34   0.733    -.0914573    .0643837
------------------------------------------------------------------------------
```