

Department of Economics

Tests of Utility Independence When Health Varies over Time

Anne Spencer and Angela Robinson

Working Paper No. 596

May 2007

ISSN 1473-0278



Tests of utility independence when health varies over time

Anne Spencer ¹ & Angela Robinson ²

¹ Department of Economics, Queen Mary, University of London

² School of Medicine, Health Policy & Practice, University of East Anglia

May 2007

Abstract

In the conventional QALY model, people's preferences are assumed to satisfy utility independence. When health varies over time, utility independence implies that the value attached to a health state is independent of the health state that arise before or after it. In this paper we set out to test the extent to which utility independence is undermined by sequence and duration effects. Two separate studies were conducted involving a total of 155 respondents. In study one, we conducted 5 tests of utility independence using a standard gamble question. Three of the tests of utility independence were repeated in study two after randomisation was introduced in order to take account of possible ordering effects. Utility independence holds in the majority of cases examined here and so our work generally supports the use of utility independence to derive more tractable models.

KEY WORDS: utility independence, QALY.

JEL classification: H5, I10.

Acknowledgements

We would like to thank participants at the Preference Elicitation Group in London in December 2003 where this study was initially presented. Thanks also goes to Stella Kaltjob who discussed our paper at CES/HESG in Paris in January 2004. We would also like to thank Graham Loomes and Peep Stalmeier for comments on the design and Judith Covey for statistical advice on the analysis of order effects. We are also very grateful for the detailed comments given by two anonymous referees. Financial support from the Medical Research Council is gratefully acknowledged .

1. Introduction

Governments are increasingly drawing upon survey techniques to incorporate people's opinions into policy, the interpretation of which often involves a set of simplifying assumptions drawn from economic theory. We investigate the application of Keeney and Raiffa's (1976) utility independence assumption to interpret people's preferences towards risky treatments when health varies over time (Drummond et al, 1997). The descriptive validity of utility independence has been questioned by psychologists who argue that people may have preferences over the sequencing and duration of health states (Loewenstein and Prelec, 1993). Such doubts have led to the use of alternative means of incorporating preferences that relax somewhat the assumption of utility independence (Guerrero and Herrero ,2005).

In this paper we set out to test utility independence when health states vary over time, which, as far as we are aware, has not been previously tested. The Background reviews other authors' tests of independence for both chronic health states and when health varies over time and provides the motivation for the current study. In the Method and Results, we outline the design and main findings. These findings are further reviewed in the Discussion and we suggest ways for future research in the Conclusion.

2. Background

Quality adjusted life years (QALYs) are used by health economists to quantify people's preferences towards treatments that differ in terms of quality of life and life expectancy. When health states vary over time the QALY approach assumes that it is valid to estimate the utility of the health states independently of one another and simply adds these utilities

(appropriately weighted by a measure of a respondent's preferences for time). For example, suppose a person experiences health state X in period 1, Y in period 2 and Z in period 3, and we represent this by the health profile XYZ. The QALY approach then estimates the utility of the profile XYZ using equation (1).

$$U(XYZ) = w_1 U_1(X) + w_2 U_2(Y) + w_3 U_3(Z) \quad (1)$$

where w_i is the time discount factor and $U_i(.)$ is the utility function at time i , for $i=1,2,3$.

When health states vary over time, Bleichrodt (1995) and Bleichrodt & Quiggin (1997) show that for QALYs to be a valid measure under Expected Utility Theory it is necessary to assume additive independence over disjoint time periods. Additive independence holds if the preferences between risky treatments depend only upon the marginal rather than the joint probability distributions of the health states (Bleichrodt & Quiggin, 1997, p.154; Keeney & Raiffa, 1976, p.230 and p.263, Fishburn 1965). Under additive independence a respondent should be indifferent between a risky treatment with a 0.5 chance of profile XXX and 0.5 chance of YYY and a risky treatment with a 0.5 chance of YXX and 0.5 chance of XYY. In this example, we have underlined the states that are varied in the test. There is limited evidence on additive independence, but Spencer (2003) observed some violations of additive independence.

Additive independence is strong assumption and may not always hold. But this does not imply an end to the QALY approach, because if the weaker assumption of utility independence still holds then this can be used as the basis to derive models that are more tractable. If so, these model may not be as simple as the conventional QALY model and may require the estimation of weights for different phases of a person's life cycle (see page 33 Bleichrodt, 1995). There are two lines of investigation of utility independence: one for

chronic health states the other for health states that vary over time. Research suggests that utility independence holds for chronic health states, although the tests of utility independence for chronic states are rather different to those for the when health states vary over time. Miyamoto and Eraker (1988) found that a respondent's risk attitude towards different survival durations was unaffected by health quality and concluded that survival duration is utility independent of health quality for chronic health states. Bleichrodt and Johannesson (1997) found that utility scores are unaffected by duration after allowing for the imprecision of preferences and concluded that quality was utility independence of survival duration for chronic states. Doctor et al. (2004) also found support for the QALY model for chronic states, and by implication utility independence for chronic states. Finally, Bleichrodt and Pinto (2005), also found support for utility independence for chronic states in a model that took account of violations of Expected Utility.

Much less is known about the situation where health states vary over time. When health states vary over time utility independence holds if preferences between risky treatments, that contain the same health state in period i do not depend upon the severity of the health state in period i (Keeney & Raiffa, 1976, p.226). However, this utility independence assumption does not appear to have been tested which we consider to be an important omission from the literature and we aim to address in our paper. Treadwell (1998) tested a special case of utility independence, where all alternatives are certain, termed preferential independence. Preferential independence was satisfied in the majority of cases (36 out of the 42 tests) even though his tests were designed to be sensitive to violations of preferential independence, Further, preferential independence held regardless of the discount rate used. However, preferential independence is not sufficient to imply useful models when alternatives are risky.

There is reason to doubt that utility independence will hold when health varies over time. Psychologists have argued that respondents may have preferences over the ordering of events, known also as sequencing effects (Gafni, 1995; Ross & Simonson, 1991). A respondent may experience ‘dread’ and desire to overcome ill-health in the short term or prefer to ‘savour’ the prospect of good health in the long term (Loewenstein & Prelec, 1993). A respondent may also pay more attention to the final health state in a profile (Kahneman, et al 1993; Varey & Kahneman 1992) and under-weight earlier states. They may also adapt to health in a positive or negative manner over time (Ross & Simonson, 1991).

Guerrero and Herrero (2005) recently relaxed utility independence in a semi-separable QALY approach that allows for some sequence effects (see also Meyer Chapter 9 in Keeney and Raiffa, 1976). In so doing, they distinguish between ‘initial independence’ and ‘final independence’. In the former, the conditional preferences for lotteries over the *final* health states are independent of the *initial* health states. In the latter, preferences for lotteries over the *initial* health states are independent of the *final* health states. The semi-separable QALY approach requires only that ‘initial independence’ applies. The approach can, therefore, incorporate respondents’ preferences for increasing or decreasing profiles over time. In addition, the model can allow for duration effects, whereby prolonged exposure to severe states lead to a decrease in utility. However, no information exists on the descriptive force of their conditions which is something we aim to do also in this paper.

Thus, we set out to design a study that tested utility independence and the extent to which initial and final independence holds. The aims and objectives of the study are:

To carry out a test of utility independence in risky choices for mild and severe health states.

To test the impact of changing health at the beginning or end of a health profile.

3. Methods

Four states were used in these profiles and were colour-coded such that normal health (N) was represented by pink, mild disability (Y) by yellow, severe disability (B) by blue, and death (D) by black. The health state descriptions were taken from a Health and Safety financed project which investigated the impact of health states in the long term and are given in figure 1.

A set of ‘life profiles’ were developed each covering a 25 years period, made up of 5 periods of 5 years. In the notation that we use NNNBB denotes 15 years in normal (N), followed by 10 in the severe disability (B), whereas YNNNN denotes 5 years in the mild disability (Y) followed by 20 years in normal health.

The test of utility independence was based on a SG question that it is commonly used in health economics to elicit utilities of health states¹. Respondents were first asked a ‘practice’ SG question in order to familiarise them with the response format. This question asked them

¹ Keeney and Raiffa’s test of utility independence in the QALY model would involve setting the probability of the risky treatment at 0.5, and asking respondents to vary outcomes in the certain treatment to compensate for variations in the severity of health in period \bar{i} . We considered that it would be difficult to ask respondents to vary the severity of health in this manner. In contrast, we set the outcomes and asked respondents to vary the probability of the risky treatment to compensate for variations in the severity of health in period \bar{i} . Adapting the question in this way allowed us to base the test on a SG question that it is commonly used in health economics to elicit utilities of health states.

to compare a risky treatment with NNNNN (25 years in normal health) as the best outcome and DDDDD (death 25 years early) as the worst outcome, to the certainty of NNDDD (10 years in normal health followed by death). Respondents were then presented with a table that showed the chances of success and failure associated with the risky treatment. For each chance of success and failure they were asked to consider whether they preferred the certainty, preferred the risky treatment, or found it too hard to choose between those two options and the mid-point value was taken between the choices they found hard to choose). Respondents were encouraged to consider the top and bottom of the table first and state their preference and to then work through the rest of the rows at their own speed.

After completing the practice question, respondents were presented with five tests of utility independence, each test comprising of two SG questions, A and B, making ten SG questions in all. Table 1 details the five tests of utility independence. In this table we underline the states that are varied between the two 'halves' of each test. For example, in question 5A respondents were offered the certainty of profile NNNBN and a p_A chance of profile NNNNN and $1-p_A$ chance of BBBBN. Whereas in question 5B they were offered the certainty of profile NNNBB and a p_B chance of profile NNNNB and $1-p_B$ chance of BBBBB. For each question the chances of success and failure were given in a table in a SG booklet. The null hypothesis is that utility independence holds and $p_A = p_B$. In each case, the two 'halves' of the independence test were answered consecutively. This was done in order to minimise the possibility that any differences detected between the two treatments were due to 'random noise' by encouraging respondents to compare parts A and B directly.

Tests 1 and 5 explore the impact of changing the health state in the last period from normal health to the severe disability. Evidence that preferences are affected by the final health states in these tests would be consistent with the semi-separate QALY approach of Guerrero and Herrero (2005) but not the conventional QALY model. Tests 1 and 5 also allows an examination of the extent to which responses are consistent between questions since the best and worst outcomes associated with these risky treatments are identical. For consistency we would expect the chance at which a respondent is indifferent in 1A to be lower than in 5A, and that in 1B to be lower than in 5B, since the certain outcome is worse in 1A compared to 5A, and worse in 1B compared to 5B. Tests 2 and 3 explore the impact of changing the health state in the first period from normal health to severe disability (test 2) or to mild disability (test 3). It is important to note that both Guerrero and Herrero's (2005) semi-separable QALY approach and the conventional QALY are undermined if changes in the initial health states affect the preferences for the final health states².

Test 4 examines the impact of changing the health state in the first and second period from normal health to severe disability. Test 4 differs from all the other tests in two main ways: a) health changes for 10 years across the two 'halves' of the test and b) there is a prospect of premature death in the worst outcome. It seems plausible that both serve to increase the salience of the severity of the disability, making violations of independence more likely.

We first carried out a series of paired sample t-tests, in each case comparing responses to part A and part B, using a significance level of 5%. We corrected for repeated testing by reducing

² Guerrero and Herrero's model is a dynamic decision model where preferences are independent of what happened in the past. To test initial independence fully in their model would require asking questions at different points in time.

the significance level of each test following the procedure suggested by Bonferroni (Maxwell and Delaney, 1990). The significance level was reduced to 0.01 in study one (i.e. $0.05/5=0.01$, where there were five comparisons) and to 0.0167 in study 2 (i.e. $0.05/3=0.0167$, where there were three comparisons).

A convenience sample of students registered at the Economics Department at Queen Mary University of London was used. Students who took part in the study were naive to the hypothesis being tested and to methods used to measure health state utility. Between 10 and 20 students took part in each session. The researcher illustrated the SG questions using overhead slides and checked that respondents understood the practice question. The students were then asked to progress through the SG booklet at their own speed.

4. Results of study one

The sample comprised of 64 respondents, 37 males, 27 females with a mean age of 21.

Table 2a shows the mean, median and standard deviations of the responses, given in terms of the chance of success where respondents were indifferent between the two treatments. For consistency we would expect that the chance at which a respondent is indifferent in 1A to be lower than in 5A, and that in 1B to be lower than in 5B. Of the 64 respondents, 42 were strictly consistent in both tests (i.e. $p_{1A} < p_{5A}$ and $p_{1B} < p_{5B}$, where the subscripts denote both the question and part), 2 were strictly consistent in one test and weakly consistent in the other (i.e. $p_{1A} > p_{5A}$ and $p_{1B} = p_{5B}$ or vice versa), 14 respondents were strictly inconsistent in at least one test (i.e. $p_{1A} > p_{5A}$ and/or $p_{1B} > p_{5B}$), whilst the pattern was indeterminate for the remaining 6 respondents due to missing data.

The results of paired t tests comparing the two ‘halves’ A and B of each test are given in table 2b. Clearly, there is no significant difference at the 5% level between responses to parts A and B in the case of four of the five independence tests carried out. The data were also re-examined after the 14 respondents giving at least one strictly inconsistent response had been removed with no change in the results. For each of the 5 tests in turn, the p values without (with) ‘inconsistent’ respondents were as follows: 1 (0.494) for test 1, 0.493 (0.479) for test 2, 0.311 (0.155) for test 3, 0.285 (0.864) for test 4 and 0.000 (0.001) for test 5. Hence, we have to conclude that utility independence generally holds in the way we set out to examine it here.

It is only in test 5 that we do find a significant difference, in particular, a significantly greater number of respondents set the indifference value of p higher in question 5A than in 5B ($p_A > p_B$). Further, the null hypothesis of independence for test 5 is still rejected after the significance level is adjusted in order to allow for repeated tests (in this case, to $p=0.01$). This finding is slightly puzzling as test 5 was identical to test 1 other than the duration of severe health under the certain outcome.

5. Caveats to study one

In study one, utility independence was found to hold in general although we did find a significant difference in the case of one of the tests - namely test 5 - whereby respondents were significantly more likely to set p higher in part A than part B. We identified a number of caveats to the tests carried out in study one. First, it seemed plausible that a sample of economic students, even with no specific knowledge of the QALY model and naïve to the hypothesis under examination, may have a desire to give responses that are apparently

‘consistent’ or ‘rational’. By presenting the two halves of the test consecutively in each case may have made it more likely that they identified what the ‘rational’ response was in each case. Further, we did not randomise the order of the parts to each test, with part A – the ‘better’ half of the test – always preceding part B. This may have made the data susceptible to anchoring and adjustment effects, the impact of which being hard to predict. More importantly perhaps, we failed to randomise the order of the tests themselves, with test 5 always appearing last in the response booklet. It is impossible, therefore, to say whether the pattern uncovered between tests 1 and 5 (which were identical except for period of severe health in the certain outcome) was some sort of ordering effect.

6. Study two

A second study was conducted using the same sampling procedures as before, and drawing respondents from a convenience sample of students registered at the Economics Department at Queen Mary University of London. As far as possible the procedures were identical to those described in study one, again with groups of between 10-20 respondents taking part in the study. The crucial difference between the two studies being that, in study two, the tests (and ‘halves’ of the tests) were presented to respondents in random order and so each respondent received a booklet containing a different ordering of questions. Due to time constraints, we elected to repeat only three of the five tests carried out previously. These are shown in Table 1 as tests 1, 4 & 5. Thus, in study two respondents were presented with a series of six SG questions in a random order (i.e. parts A and B of tests 1, 4 & 5 respectively)

³. The health state descriptors, visual stimuli, response sheets and verbal instructions were identical to those used in study one.

7. Results of study two

The sample for study two comprised of 92 respondents, 48 males, 44 females with a mean age of 20. Table 3a shows the mean, median and standard deviations of the responses, again given in terms of the chance of success where respondents were indifferent between the two treatments. Again, we would expect that the chance at which a respondent is indifferent in 1A to be lower than in 5A, and that in 1B to be lower than in 5B. Of the 92 respondents, 51 were strictly consistent in both tests (i.e. $p_{1A} < p_{5A}$ and $p_{1B} < p_{5B}$, where the subscripts denote both the question and part), 15 were strictly consistent in one test and weakly consistent in the other (i.e. $p_{1A} < p_{5A}$ and $p_{1B} = p_{5B}$ or vice versa) 2 were weakly consistent in both tests (i.e. $p_{1A} = p_{5A}$ and $p_{1B} = p_{5B}$), 20 respondents were strictly inconsistent in at least one test (i.e. $p_{1A} > p_{5A}$ and/or $p_{1B} > p_{5B}$), whilst the pattern was indeterminate for the remaining 4 respondents due to missing data.

The results of paired t tests comparing the two ‘halves’ A and B of each test are given in table 3b. There is no significant difference at the 5% level between responses to parts A and B in the case of tests 4 and 5. Again, the data were re-examined after the 20 ‘inconsistent’ respondents had been removed (those respondents who set the indifference value of p to be higher in test 1A than 5A and/or higher in 1B than 5B) with no significant change in results.

³ A random numbers generator was used to generate a random ordering of the 6 questions faced by respondents in study two.

For each of the three tests in turn, the p values without (with) ‘inconsistent’ respondents were as follows; 0.016 (0.008) for test one, 0.690 (0.973) for test four and 0.117 (0.417) for test five.

Contrary to the results in study one, it is only in test 1 that we now find a significant difference, in particular, a significantly greater number of respondents set the indifference value of p higher in question 1A than in 1B ($p_A > p_B$). Further, the null hypothesis of independence for test 1 is still rejected after the significance level is adjusted in order to allow for repeated tests (in this case, to $p=0.0167$ as 3 comparisons were made). This finding was unexpected and is the opposite effect to which we found in the first study (recall that in study one test 5 was significant, whilst test 1 was not).

8. Discussion

Treadwell (1998) tests preferential independence when health varies over time under conditions of certainty. We set out to see whether independence holds under conditions of uncertainty, termed ‘utility independence’, and our tests are a valid test of utility independence under both Expected Utility and Non-Expected Utility models (Miyamoto & Wakker, 1996). Whilst our results were somewhat mixed, we find that independence generally holds in the way that we set out to examine it here. It has been shown elsewhere that, if utility independence holds then it is still possible to derive models that are tractable, even if the stronger assumption of additive independence fails (Bleichrodt, 1995). On the face of it then, our findings are generally supportive of the use of QALYs in health care decision making, providing that an appropriate specification of the model is used (see Bleichrodt, 1995 for details).

This finding, however, runs contrary to the evidence cited that sequence and duration effects *do* matter to people (Ross & Simonson, 1991; Loewenstein & Prelec, 1993; Kahneman, at al 1993; Varey & Kahneman 1992) and we consider below the possible reasons for this discrepancy . First, the task respondents undertook is fairly complex and they may have adopted ‘simplifying strategies’ in order to get through it. For example, respondents may have ‘edited out’ information that was obviously common across choices in order to simplify the task, making violations of independence less likely. Thus, if it was obvious to respondents that one period of the profile was common across the risky and certain outcomes, the severity of this common period may have been ignored altogether- guaranteeing utility independence as assessed here.

Second, the task presented to the respondents was highly abstract and it is possible that the stimulus failed to adequately emphasize the *changing* patterns of health status over time. It is possible that using stimuli that are more *dynamic* in nature, may better represent changes in health over time than the ‘static’ representations used here. For example, Chapman (2000) used graphs to depict *changes* in the quality of health over time, which are arguably more appropriate stimuli to test for sequencing and duration effects than those used here. This is an issue that may be addressed in further studies.

Certain of our findings, however, are more difficult to explain. Whilst it seemed plausible that independence would be less likely to hold when the tests were presented in random order there is no obvious reason why test 1 should be significant in study two whilst test 5 is not (reversing the previous finding). We believe that more weight ought to be placed on the findings of the second of the two studies that controlled for ordering effects, but this remains a finding to be investigated further.

One way forward for future research would be to conduct a qualitative study alongside the quantitative tests to reveal those factors that were considered by respondents in formulating their responses. Such data may help determine whether there is a psychological explanation for those violations of independence that were uncovered here or whether they were an artefact of the study design.

More generally, it may be argued that asking respondents to consider the type of stylised, hypothetical scenarios used here, will necessarily fail to capture feelings of adaptation, savouring and dread that may matter to people in real life. Kahneman and Sugden (2005) distinguish between *experienced* utility (utility as hedonic experience) and *decision* utility (utility as a representation of preference) and argue that the latter will underestimate feelings of adaptation etc. Others may argue that if decision utility corresponds more closely to what people think should influence their choices and is more normative in nature, then decision utility is more relevant for economic evaluations that are principally normative in character. Whilst it is not clear to us that we could (or even should⁴) abandon the practice of eliciting values over hypothetical health states in utility assessment exercises, there may be particular problems inherent in attempting to assess preferences over *sequences* of health states.

9. Conclusions

We set out to see whether independence holds under conditions of uncertainty termed ‘utility independence’. We find that utility independence holds in the majority of cases examined

⁴ As this would necessarily rule out the use of general population values for health states in favour of those of disparate groups of patients, making comparability and aggregation problematic.

here. In particular, changing the health state at either the beginning or the end of a profile did not have a significant impact on preferences over the remainder of that profile.

Even if further research supports the currently limited evidence that additive independence fails (Spencer, 2003), then utility independence can be used as the basis to derive models that are tractable and can be applied to practical research. Whilst these models are unlikely to be as simple as the *conventional* QALY model, it does not sound the death knell for the use of QALYs in economic evaluation.

References

Bleichrodt, H. (1995). QALYS & HYE: Under what conditions are they equivalent? *Journal of Health Economics*, 14, 17-37.

Bleichrodt, H. & Pinto J. (2005). The validity of QALYs under non-expected utility, *Economic Journal*, 115, 533-550.

Bleichrodt, H., & Quiggin, J. (1997). Characterizing QALYs under a general rank dependent utility model. *Journal of Risk of Uncertainty*, 15, 151-165.

Bleichrodt, H., & Johannesson, M. (1997). An experimental test of constant proportional tradeoff and utility independence, *Medical Decision Making*, 17, 21-32.

Chapman, G.B. (2000). Preferences for improving and declining sequences of health outcomes *Journal of Behavioral Decision Making*, 13 (2): 203-218.

Doctor, J.N., Bleichrodt, H., Miyamoto, J.M., Temkin, N.R., & Dikmen, S. (2004). A new and more robust test of QALYs. *Journal of Health Economics*, 23, 353-367.

Drummond, M.F., O'Brien, B., Stoddart, G.L. and Torrance, G.W. (1997) *Methods for the economic evaluation of health care programmes*. second edition, Oxford: Oxford University Press.

Fishburn, P.C. (1965) Independence in Utility Theory with whole product sets, *Operations Research*, 13, 28-45.

Guerrero, A.M., & Herrero, C. (2005) A semi-separable utility function for health profiles. *Journal of Health Economics*, 24, 35-54.

Gafni, A. (1995). Time in health: can we measure individuals' "pure time preference". *Medical Decision Making*, 15, 31-37.

Kahneman, D., Fredrickson, B.L., Schreibern, C.A., & Redelmeier, D.A. (1993). When more pain is preferred to less: adding a better end. *Psychological Science*, 4, 401-405.

Kahneman, D. and Sugden, R, (2005), Experienced utility as a standard of policy evaluation, *Environmental & Resource Economics*, 32 (1): 161-181.

Keeney, R.L., & Raiffa, H. (1976). *Decisions with multiple objectives, preferences and value tradeoffs*. London: Wiley.

Loewenstein, G., & Prelec, D. (1993). Preferences for sequences of outcomes. *Psychological Review*, 100, 91-108.

Maxwell, S., & Delaney, H. (1990). *Designing experiments and analyzing data*. Belmont, C.A: Wadworth.

Miyamoto, J.M., & Eraker, S.A. (1988). A multiplicative model of utility of survival duration and health quality, *Journal of Experimental Psychology: General*, 117, 3-20.

Miyamoto JM & Wakker,P (1996), Multiattribute utility theory without Expected utility foundations, *Operations Research*, 44: 313-326

Ross, W.T., & Simonson, I. (1991). Evaluating pairs of experiences: a preference for happy endings. *Journal of Behavioral Decision Making*, 4, 273-282.

Spencer, A. (2003). A test of the QALY model when health states varies over time, *Social Science and Medicine*, 57, 1697-1706.

Treadwell, J.R. (1998). Tests of preferential independence in the QALY model. *Medical Decision Making*, 18, 418-428.

Varey, C., & Kahneman, D. (1992). Experiences extended across time: evaluation of moments and episodes. *Journal of Behavioral Decision Making*, 5, 169-185.

Table 1: The five tests of utility independence

		Better outcome, p	Worse outcome,(1-p)	Certainty of:
Test one	1A	<u>NNNN</u>	<u>BBBB</u>	<u>NNBB</u>
	1B	<u>NNNB</u>	<u>BBBB</u>	<u>NNBB</u>
Test two	2A	<u>NNNN</u>	<u>NBBBB</u>	<u>NNBB</u>
	2B	<u>BNNNN</u>	<u>BBBBB</u>	<u>BNNBB</u>
Test three	3A	<u>NNNN</u>	<u>NYYYY</u>	<u>NNYY</u>
	3B	<u>YNNNN</u>	<u>YYYYY</u>	<u>YNNYY</u>
Test four	4A	<u>NNNN</u>	<u>NNDDD</u>	<u>NNYY</u>
	4B	<u>BBNNN</u>	<u>BBDDD</u>	<u>BBYY</u>
Test five	5A	<u>NNNN</u>	<u>BBBB</u>	<u>NNBN</u>
	5B	<u>NNNB</u>	<u>BBBB</u>	<u>NNBB</u>

Table 2a: Responses to the five tests of independence in Study One

	Part	N	Mean (Std)	Median
Test 1	A	62	0.601 (.165)	0.575
	B	62	0.619 (.190)	0.600
Test 2	A	63	0.622 (.165)	0.625
	B	63	0.606 (.197)	0.625
Test 3	A	63	0.555 (.185)	0.550
	B	63	0.596 (.169)	0.575
Test 4	A	63	0.715 (.191)	0.725
	B	62	0.716 (.189)	0.750
Test 5	A	62	0.774 (.186)	0.800
	B	61	0.731 (.172)	0.725

Table 2b: Results of paired t tests in Study One			
Pair-wise comparison	Mean difference	Standard error	P value
1A – 1B	-0.0216	0.016	0.494
2A – 2B	+0.0018	0.020	0.479
3A – 3B	-0.0348	0.025	0.155
4A – 4B	+0.0027	0.017	0.864
5A – 5B	+0.0497*	0.014	0.001

*the mean difference is significant at 0.05 level

Table 3a: Responses to the three tests of independence in Study Two

	Part	N	Mean (Std)	Median
Test 1	A	87	0.603 (.189)	0.575
	B	87	0.553 (.194)	0.525
Test 4	A	86	0.674 (.205)	0.700
	B	86	0.675 (.218)	0.675
Test 5	A	90	0.706 (.189)	0.725
	B	90	0.692 (.188)	0.675

Table 3b: Results of paired t tests in Study Two			
Pair-wise comparison	Mean difference	Standard error	P value
1A – 1B	+0.0505*	0.019	0.008
4A – 4B	-0.0067	0.019	0.973
5A – 5B	+0.0137	0.017	0.417

*the mean difference is significant at 0.05 level

Figure 1. The health states

Pink health state	Yellow health state	Blue health state
<p>Compared to the average person that age;</p> <ul style="list-style-type: none"> • Your indoor activities are not limited. • Your outdoor activities are not limited. • You are full of energy. 	<p>Compared to the average person that age;</p> <ul style="list-style-type: none"> • Your indoor activities are <u>a bit more</u> limited. • Your outdoor activities are <u>a bit more</u> limited. • You have <u>a bit less</u> energy. 	<p>Compared to the average person that age;</p> <ul style="list-style-type: none"> • Your indoor activities are <u>a lot more</u> limited. • Your outdoor activities are <u>a lot more</u> limited. • You have <u>a lot less</u> energy.

**This working paper has been produced by
the Department of Economics at
Queen Mary, University of London**

**Copyright © 2007 Anne Spencer and Angela Robinson
All rights reserved**

**Department of Economics
Queen Mary, University of London
Mile End Road
London E1 4NS
Tel: +44 (0)20 7882 5096
Fax: +44 (0)20 8983 3580
Web: www.econ.qmul.ac.uk/papers/wp.htm**