



Universiteit
Leiden
The Netherlands

The construction of health state utilities

Osch, S.M.C. van

Citation

Osch, S. M. C. van. (2007, September 6). *The construction of health state utilities*. Retrieved from <https://hdl.handle.net/1887/12363>

Version: Corrected Publisher's Version

License: [Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden](#)

Downloaded from: <https://hdl.handle.net/1887/12363>

Note: To cite this publication please use the final published version (if applicable).

4

The construction of standard gamble utilities

The construction of standard gamble utilities.

S.M.C. van Osch, A.M. Stiggelbout

Health Economics: 2007; in press.

Abstract

Health effects for cost-effectiveness analysis are best measured in life years, with quality of life in each life year expressed in terms of utilities. The standard gamble (SG) has been the gold standard for utility measurement. However, the biases of probability weighting, loss aversion and scale compatibility have an inconclusive effect on SG utilities. We determined their effect on SG utilities using qualitative data to assess the reference point and the focus of attention. While thinking aloud, 45 healthy respondents provided SG utilities for six rheumatoid arthritis health states. Reference points, goals, and focuses of attention were coded. To assess the effect of scale compatibility, correlations were assessed between focus of attention and mean utility. The certain outcome served most frequently as reference point, and the SG was perceived as a mixed gamble. Goals were mostly related to this outcome. Scale compatibility led to a significant upward bias in utilities; attention was relatively more with the low outcome and this was positively correlated with mean utility. SG utilities should be corrected for loss aversion and probability weighting with the mixed correction formula proposed by prospect theory. Scale compatibility will still bias SG utilities, calling for research on ways to correct for this bias.

Introduction

In cost-effectiveness analyses, the costs associated with a health care intervention are compared to its benefits (health effects). It is commonly acknowledged that health effects are best measured in terms life years and that quality of life in each life year is best expressed in terms of utilities, that are then used as a weighting factor, yielding quality-adjusted life years (QALYs) (8;9;48). Based on normative expected-utility arguments, the standard gamble (SG) method has often been considered the gold standard for utility measurement because, unlike other elicitation methods, it incorporates risk. The standard gamble generally requires a respondent to compare the certainty of being in the health state to be valued for the remaining life expectancy, with a gamble that offers a chance of optimal health for the remaining life expectancy but also entails a risk of immediate death. In the generally used probability equivalent of the SG, the respondent is asked to indicate at what probabilities of the gamble he or she would be indifferent to the choice between the health state and the gamble.

There is much empirical evidence demonstrating that expected utility is not descriptively valid, the three main reasons for this being probability weighting, loss aversion, and scale compatibility (6;7;11;13). These biases generally lead to SG utilities that are too high. The use of biased utilities may lead to biased resource allocation decisions and, therefore the joint effect of these biases should be minimized in the elicitation of utilities. This leaves a great need for more knowledge about these biases in health utility measurement, so as to adequately correct for them or to be aware of their effect on SG utilities (and health care choices) (4).

We will first discuss the biases of loss aversion and probability weighting, because they interact. Loss aversion implies that the disutility that a person experiences from

losses is significantly greater than the utility the person experiences from gains of the same absolute size (7). Losses weigh more heavily in decisions than gains do. Probability weighting entails that people weight probabilities in a nonlinear manner. Probability weighting will usually lead to upward bias for SG utilities (6). Cumulative prospect theory (PT) proposes to transform probabilities by using a rank-dependent probability-weighting function (5;49).

The weighting function that corrects for probability weighting and loss aversion depends on the perception of the gamble. The SG can be perceived as yielding all losses, all gains, or as being mixed (yielding both gains and losses), depending on the perceived reference point. SG utilities are only distorted by loss aversion if the SG is perceived as a mixed gamble. To adequately correct for probability weighting, or to know whether loss aversion influences SG utilities, knowledge about the location of the reference point is needed. However, PT incorporates no theory with respect to the reference point, and direct evidence for the location of the reference point is absent.

Hershey and Shoemaker, as well as Bleichrodt et al., reasoned that the certain outcome is fixed in SG elicitation and, therefore, may provide a salient reference point (6;7). Robinson et al. argued that a task instruction could lead to adopting a certain reference point (50). For example, if the instruction reads that respondents should imagine themselves 'to be' in the health state to be valued, i.e. the certain outcome, after which a medical intervention is to be followed that can result in success (a return to optimal health) or failure (death), this will probably cause respondents to adopt the certain outcome as the reference point. Robinson et al. indeed observed this instruction effect on the reference point in their study. The scale-compatibility bias results from the principle that the attributes of decision alternatives that are compatible with the response scale are weighted more heavily in decisions (6;23).

In the probability equivalent version of the SG, the response scale is a probability, and, thus, the individual will focus on the probability. The associated outcome will receive more weight. The problem is, as Bleichrodt argued, that the bias may hold equally well for the certain outcome probability, the good-outcome probability as for the bad-outcome probability (6). Scale compatibility leads to higher utilities if a respondent focuses on the outcome that involves the bad-outcome probability. The gamble would become less appealing compared to the certain outcome. To achieve indifference, the gamble must be made more appealing which is accomplished by increasing the good-outcome probability. Following a similar reasoning, scale compatibility would lead to lower utilities if a respondent focuses on the good-outcome probability, as then the gamble is found to be more appealing than the certain outcome. This bias leads to a preference for the gamble over the certain outcome. Indifference is achieved by decreasing the good-outcome probability. Thus, the scale-compatibility bias could lead to either higher or lower SG utilities. It is unknown whether one of the outcomes (probabilities) draws more attention than the other. Given that the probability of the certain outcome is not varied ($p = 1$), we assume that either an upward bias on SG utilities (due to focus on the bad outcome) or a downward bias on SG utilities (due to focus on the good outcome) will occur.

The purpose of this study was to further explore SG biases using qualitative data (using a 'think-aloud' protocol). The aims were twofold. The first aim was to locate the SG outcome that is the reference point or seems to lie closest to the reference point. In the rest of the paper, we speak of 'the outcome that serves as reference point'. With knowledge of the reference point, the proper correction method can be applied to counteract probability weighting and loss aversion (the latter if necessary). In an earlier study, we observed that the desire to attain goals (e.g. raise children, make a career) probably influences the perception of the reference point in certainty

equivalent standard gambles (51). In a non-health setting, one study even argued that goals served as the reference point (44). Therefore, we focused on goals in relation to the outcomes as well.

The second objective was to obtain an indication of whether scale compatibility results in a systematic bias upwards or downwards, i.e. whether respondents focus more on the bad outcome or the good outcome? To assess this point, we identified the focus of attention. Additionally, we aimed to verify that a main focus on a bad outcome or good outcome leads to higher or lower utilities respectively. Furthermore, relevant themes raised by respondents during the experiment, that could result in a biased utility, were also taken into account.

Methods

Sample

Forty-five respondents, recruited via newspaper advertisements and pamphlets, participated in the study. Each respondent was paid € 22.50 for participation in two interviews (conducted by the first author, SMCvO); one of which is the topic of this paper. The interview took 45 minutes on average to complete. No specific sample criteria were applied.

Procedure

Six rheumatoid arthritis health state descriptions were selected from those given by rheumatic patients in the Rheumatoid Arthritis Patients In Training study (52). Descriptions were according to the EQ-5D system, a multi-attribute health-utility system. The EQ-5D system includes five dimensions of health (mobility, self-care,

usual activities, pain/discomfort, and anxiety/depression). Each dimension comprises three levels (no problems, some/moderate problems, and extreme problems). A unique EQ-5D health state combines one level from each of the five dimensions. The health states were chosen to cover the utility continuum (0 to 1), using corresponding EQ-5D valuations from the general public based on the TTO (25). We used the EQ-5D health state descriptions; 21232 (I, EQ-5D index of .09), 22322 (II, EQ-5D index of .19), 21321 (III, EQ-5D index of .36), 21222 (IV, EQ-5D index of .62), 21211 (V, EQ-5D index of .81), and 21111 (VI, EQ-5D index of .85).

The SG was written using the program Ci3 (15). All elicitations were based on the choice-bracketing search procedure, i.e. a series of ping-pong questions. In total, six SGs were elicited, one for each rheumatoid arthritis health state. The order of elicitations was randomized. The experiment started with a written explanation of RA. An oral and written explanation of the SG followed, after which two examples (practice tasks) were given. We took great care regarding the instruction in order to avoid an instruction effect on the reference point as mentioned in the introduction (see Appendix 4A). Two options were given. Option 1 was a rheumatoid arthritis health state for the respondent's remaining life expectancy (LE). Option 2 was a gamble between optimal health for LE with probability p and death within a week with probability $1-p$. Probabilities in the gamble were varied until indifference resulted. LE was based on a respondent's remaining life expectancy as derived from Dutch life tables (16). At any time during an elicitation, it was possible for respondents to take a break or check earlier answers within that elicitation and change these. Elicitations ended when respondents indicated that they valued two options equally. All respondents were instructed to think aloud during the interviews without paying attention to pronouncing grammatically correct sentences. They received two examples to practice verbalizing before each task. If a subject became silent during the

task additional instructions were given to think aloud. Thinking aloud has been reported to only affect the time needed to complete the task (40), and not the task itself. Interviews were taped and transcribed.

Coding

Coding involved initial familiarization with the qualitative data, sorting and indexing of relevant themes. These steps were iterative and not strictly consecutive. Two independent coders each coded the first half of the reports to resolve any ambiguities. Observed differences in coding were discussed and for these a consensus coding was reached. One coder coded the remaining interviews for themes deduced from this process.

A "reference point" was coded if a comparison relative to a point of view was formulated, i.e. if respondents used one of the three outcomes of the SG as a starting point to indicate the difference (gain or loss) with one outcome, or with both other outcomes. The three outcomes that could serve as the reference point were: the good outcome of the gamble, the bad outcome of the gamble, and the certain outcome of the gamble. The latter was the RA health state that was valued. We coded the bad outcome as the "reference point" if it was taken as starting point, and one of both other outcomes was/were labeled as gains relative to this bad outcome. It was also coded as "reference point" if respondents indicated that they perceived this outcome, i.e. death within a week, as the current reference situation from which the SG was perceived. For example, "I have nothing and can gain 40 years of optimal health or have rheumatism if I choose option 2". Similarly, we coded the certain outcome as "reference point" if the outcome was taken as the starting point, e.g., "Either I continue living with these complaints, or I will have an operation and the operation will succeed, or it will not and I will die." We coded the good outcome as the "reference

point" if that outcome was taken as the starting point, e.g., "I will stay healthy, or lose it all, or end up with rheumatism."

A "goal" was coded if a statement was made regarding the realization of a goal with respect to an outcome. In other words, if the respondent referred to an outcome and assessed that outcome to be sufficient or insufficient for the realization of the goal.

A "focus of attention" was coded when a respondent mentioned the good or bad outcome including its probability ("There is a 50% chance to die within a week."). Additionally, "focus of attention" was coded when the good or bad outcome, including its probability, was compared to another outcome ("There is a 50% chance to die within a week and a 50% chance to live for 40 years."). There is no point of view identifiable in this comparison. We assumed that the more frequently an outcome was mentioned or compared, the more attention a respondent paid to that outcome.

Data analysis

Values of 0 and 1 were assigned to 'death within a week' and 'optimal health' respectively, and the probability of the good outcome at the point of indifference was then taken as the utility value of the RA health state. The frequency of the coded reference points was assessed across respondents. The frequency of the coded goals was assessed across respondents. We determined the relative focus of attention for each respondent by calculating whether the focus was relatively more with the good or with the bad outcome of the gamble ($\text{Total focus on bad outcome} / (\text{Total focus on bad outcome} + \text{Total focus on good outcome})$). The effect of scale compatibility on mean SG utilities was tested by calculating Pearson's R correlations between the relative focus of attention and the total mean utility for the six health states.

Results

The respondents consisted of 26 women aged 18 to 72 years (mean age = 27, s.d. = 12) and 19 men aged 19 to 61 years (mean age = 34, s.d. = 14). All respondents had received at least a high-school level of education. About 50% of the respondents were university students (mostly in biomedical science or medicine), and 25% of the respondents had children. The aggregate utility values for health states were; mean utility for health state I = .54 (s.d. = .28), mean II = .57 (s.d. = .29), mean III = .69 (s.d. = .26), mean IV = .72 (s.d. = .24), mean V = .84, (s.d. = .21), mean VI = .87 (s.d. = .17). Gender, age and the health state that is valued identify the quotes which we use to illustrate our findings. It was difficult for some respondents to combine verbalization with the task, as was apparent from the reticence of several respondents to verbalize during the task.

Reference point

The certain outcome (RA) most often served as reference point (52% of reference point codings), e.g.

"It is quite troublesome that I am not able to perform my daily activities .. I would choose for the operation so to say ... I know I will be handicapped for life if I cannot perform my daily activities. And I will have a 10% chance that I will be able to do that." (male, 61, II).

This was closely followed by the high outcome (LE) (45.5% of reference point codings), e.g.

"40% Chance to live. I do not like to gamble with my life, but I must be a good life. Yes, I think it would be equally bad to have a 60% chance to die within a week or to lose my independence" (female, 22, II).

The bad outcome (death within a week) hardly ever served as reference point (2.5 % of reference point codings), e.g.:

"A 50% to die within a week. ... Well, I die tomorrow, then I would rather live somewhat longer, but then with pain." (female, 20, IV).

Goals

Goals were mostly mentioned with respect to the certain outcome (77%), and to a lesser extent with respect to the bad outcome (14%) or the good outcome (9%). Most frequently, respondents indicated not being able to achieve their goal(s) (86%), usually with respect to the more severe health states. Most respondents mentioned more than one goal during the interview. A concern about whether goal aspiration was still possible was often expressed as a desire: 1) to either live life in a certain way (preferably in the same way as now), and/or 2) to achieve specific goals that had been set. Regarding the first concern, respondents considered the impact of the RA health state on their way of life, e.g. the ability to take part in sports, to walk, to be independent.

"What's the use of living that long without being able to do my thing?" (male, 41, III)

"A 10% chance of dying. I would certainly take the risk. To be released from the pain and the fact that I can again just do everything. That I will be able to walk." (female, 23, I)

With respect to the second concern, this was mostly related to the gamble, specifically, to the bad outcome of the gamble.

"I choose option 1. For I want the security to live, because I still have to care for my children ... My youngest is 10, and then I have a daughter who is 12, and a son who is 14 and a daughter who is 20 ... I am very involved in my children's lives." (female, 41, II)

"Yes, well, I constantly think that you are in a sort of family situation. And well, basically, it is the same argumentation over again ... Yes, well wife and children and everything ... just that you can't leave that situation and you would rather put up with the pain than take the chance of dying." (male, 23, I)

Focus of attention

Respondents barely mentioned the probability (100%) of the certain outcome. They mainly focused on the bad outcome of the gamble. Figure 4.1 presents the focus of attention per health state for the good and bad outcomes. It shows that for all health states the bad outcome drew relatively more attention than the good outcome. Additionally, focus of attention showed a significant correlation with the mean SG utility. The more the focus was with the bad outcome of the gamble the higher the mean utility was (Pearson's R correlation = .40, $p < .05$). The more the focus was with the good outcome the lower the mean utility was (Pearson's R correlation = -.40, $p < .05$).

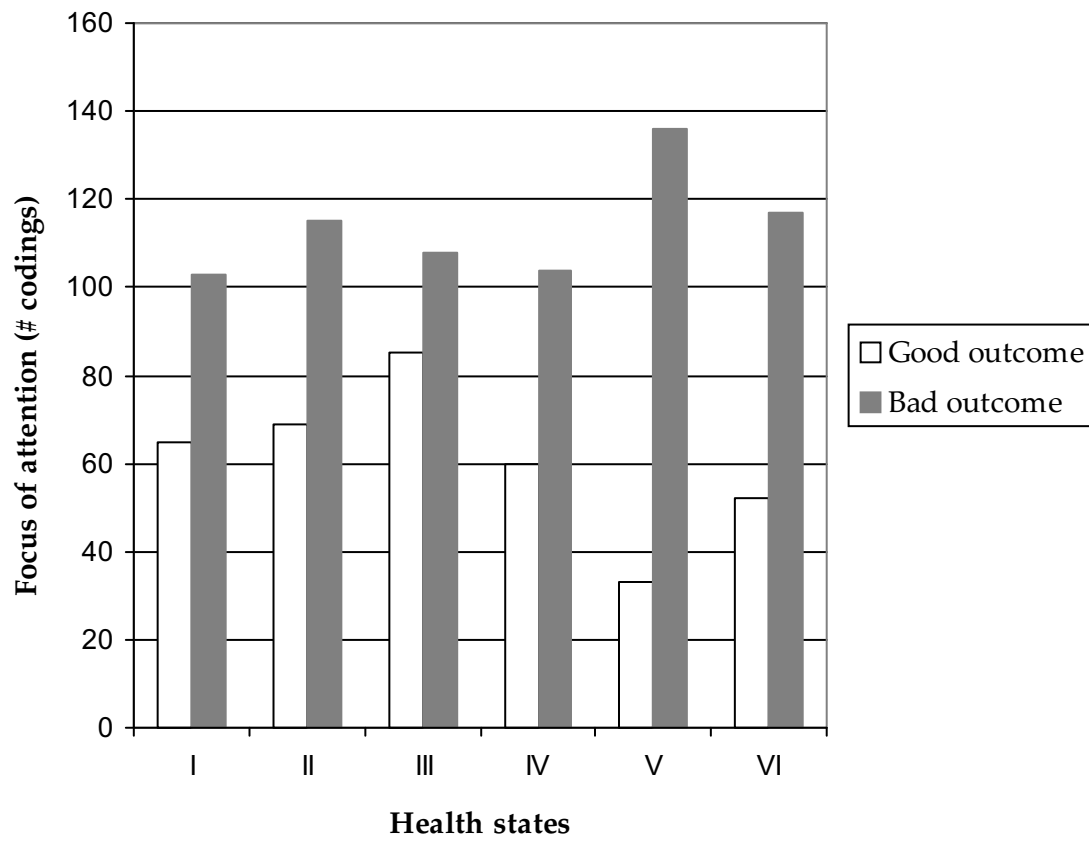


FIGURE 4.1. The frequency of focus of attention codings is depicted for the good outcome and the bad outcome of the gamble per health state. The bad outcome drew most attention.

Other themes

Other themes that may have had an effect on utilities were anticipated adaptation, and maximal endurable time (MET). Anticipated adaptation describes how respondents expected to adapt to the RA health state in due time, with a positive impact on utilities. This theme was considered fairly frequently. We distinguished between psychological and physical adaptation (53). The first is the ability to emotionally adapt to a health state, e.g. a shift in personal interests.

"I think, it may be a question of wanting ... you may think in the beginning that you cannot do it, so I, uh ... I wouldn't be up to it. To a certain point in time when you will accept it ... being so dependent, you can get used to that, I think." (female, 37, II)

Physical adaptation concerns the physical aspect of adaptation, e.g. taking painkillers.

"I don't have much knowledge about pain, but isn't there medication for that? So basically, you can live with the pain." (female, 20, I)

MET (maximal endurable time) is the concept that when people live in a health state involving a low quality of life, at first they evaluate it positively, but after a certain period (MET), any additional time spent in that health state is viewed negatively. Consequently, death is preferable after that period (54;55). METs were considered sometimes.

"I wish I could live for ten years with these problems and that it would be over after that ... to die within a week is far from attractive, but neither is going on like this for another 55 years."
(female, 26, II)

Discussion

If a person behaves perfectly in agreement with expected utility theory, then the SG method yields unbiased utility values. Empirical evidence of inconsistencies in SG valuations is well documented and has led to so-called non-expected utility theories, e.g. PT, in which biases such as probability weighting and loss aversion are described, and from which corrections can be deduced. Correcting for biases or knowledge about the effect of biases will increase the descriptive validity of the SG without sacrificing its normative validity (56). The application of qualitative techniques in our study yielded rich data on SG utilities and provided new insights into relevant biases.

Our qualitative data showed that the certain outcome most often served as the reference point. There are theoretical arguments, based on prospect theory, favoring the mixed-correction of the SG for which the certain outcome served as reference point. Another study reported the highest convergent validity between the TTO and SG if the certain outcome was used as the reference point (34). Baker and Robinson found evidence of the certain outcome serving as the reference point as well, although they did not recognize it as such. In their study, respondents framed the SG as being in the intermediate health state and choosing for the operation with a success-probability and failure-probability (57). We specifically aimed to avoid instructing respondents to adopt an outcome as the reference point. Many studies, although probably unintentionally, instruct their respondents to adopt the certain outcome as the reference point. If one wants to compare health effects between studies, as is common in cost-effectiveness analysis, it is preferable that SG utilities are elicited with the same outcome as the reference point, e.g. the health state to be valued. The biases probability weighting and loss aversion then have similar effects on SG utilities, and the proper correction method is known. An instruction may be used to this effect, in which a reference point is implied.

Not surprisingly, the low outcome of the gamble, i.e. death within a week, hardly ever served as reference point. This outcome is too far from the present health status and does not enjoy the benefits of the certain outcome. Although no theory supports the finding, the high outcome quite frequently served as the reference point. A logical explanation is that respondents involved their status quo, which likely was most equal to the high outcome, i.e. optimal health during the remaining life expectancy. After all, the situation in which the respondent has a 100% chance of having the high outcome was, most likely, their actual situation before they started the interview. This finding may be viewed as an anchoring effect.

Several goals were mentioned, mostly through concerns of not being able to achieve goals. Goals were mostly referred to with respect to the certain outcome. The relationship between goals and reference point is natural. Goals influence motivation, and thus, decision making. Lopes for example assigns a role to the aspiration level, e.g. goals, in risky decision making (58). Our findings support the hypothesis that the certain outcome most often served as the reference point.

The certain outcome probability, which does not change during the task, drew the least attention. The bad-outcome probability of the gamble appeared to draw most attention. Consequently, the theorized upwards effect of the scale-compatibility bias was observed. SG utilities were significantly higher if respondents focused more on the bad-outcome probability than on the good-outcome probability of the gamble. This provides new evidence as to why SG utilities are generally observed to be too high. It has been observed before in the medical field that people exhibit a systematic dislike for risk. Other studies report that risk avoiders generally focus on the worst outcome whereas risk seekers focus on the best outcome (59-61). Most respondents behave risk avoiding for the SG, as was the case in our study. And as expected, respondents mostly perceived the SG as a mixed gamble, and the low outcome of the gamble was therefore perceived as a loss, and over-all received most attention. Loss aversion states that a loss looms larger than a gain, which may be interpreted as support for the finding that the focus of attention lies relatively more on the low outcome.

Other themes that were relevant to our subjects were anticipated adaptation and MET. These themes may have an effect on other elicitation methods as well. The first, adaptation, may lead to higher utilities for respondents who expect to adapt to the health state. A study by Baker & Robinson reported that respondents considered

anticipated adaptation during the SG as well (57). Damschroder et al. have shown that asking respondents to perform a simple adaptation exercise, in which they were primed to anticipate their adaptation, indeed led to higher values in a person tradeoff method (62). Some respondents exhibited MET. This can lead to lower utilities, as the certain outcome is then found to be less attractive than the death outcome. This is especially so for more severely impaired health states for which maximal endurable time is likely to arise.

A limitation of our study was that approximately half of our sample consisted of university students. As a result, the findings may not generalize to the general public. A further important point is that the findings are applicable only to the choice-bracketing elicitation method. If utilities had been derived using another elicitation method (e.g. matching), these findings might have been different. Additionally, the think-aloud protocol is not without limitations. We were unable to code that which respondents did not verbalize.

Conclusions

We argue that the certain outcome, i.e. the health state to be valued, most often served as the reference point, and, therefore, a SG is most likely perceived as a mixed gamble. This can have important consequences for cost-effectiveness analyses, such as correcting preference-based utilities for loss aversion and probability weighting with the appropriate mixed correction formula (11;34). Additionally, we observed that respondents mostly focused on the low outcome of the gamble. Consequently, scale compatibility will have led to upward biases in the SG utilities.

Appendix 4A. Introduction and stimulus screen

Introduction screen

In a moment you will see an example with three possible choices. If you choose number 1, then you are choosing to definitely (100%) live for your total life expectancy with the severity of rheumatism which has been described. If you choose number 2, then you are taking a gamble: on the one hand you have the chance (...%) of living your total life expectancy in optimal health, but on the other hand, you have a chance (...%) of dying within a week. If you choose number 3, then you think that choices 1 and 2 have equal value. You don't have a preference for either choice.

Stimulus screen

Choice 1

Definitely (100%) live with rheumatoid arthritis for the rest of your life expectancy. You have:

some problems with walking
no problem washing and dressing yourself
some problems with daily activities
very severe pain or other symptoms
you are fairly fearful or gloomy.

Choice 2

A gamble between:

..% chance of living in good health for the rest of your life expectancy

OR

..% chance of dying within a week.

Choice 3

I have no preference, for me choices 1 and 2 are equal.

