DEVELOPMENT AND TEST OF A NEW METHOD FOR PREFERENCE MEASUREMENT FOR MULTISTATE HEALTH PROFILES

A Thesis
Presented to
The Academic Faculty

By

Thitima Kongnakorn

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy in
Industrial and Systems Engineering

Georgia Institute of Technology
October 2004

DEVELOPMENT AND TEST OF A NEW METHOD FOR PREFERENCE MEASUREMENT FOR MULTISTATE HEALTH PROFILES

Approved by:

Dr. François Sainfort, Advisor

Dr. Julie A. Jacko, Co-Advisor

Dr. Brani Vidakovic

Dr. Ann Bostrom

Dr. Christopher Flowers

Date Approved: October 15, 2004

ACKNOWLEDGMENTS

I am deeply indebted to my advisor, Dr. François Sainfort, and my co-advisor, Dr. Julie A. Jacko, for their constant support. Without their help, this work would not be possible. I would also like to thank the members of my committee: Dr. Brani Vidakovic, Dr. Ann Bostrom, and Dr. Christopher Flowers. Their advice and patience is truly appreciated.

I would like to thank my colleagues at Georgia Tech for their valuable suggestions and their friendship, especially to Katie Emery, Kevin Moloney, Leon Barnard, Ji Soo Yi, and Paula Edwards. Those days in HCI lab will never be forgotten. I also would like to thank Warut Tantiwong for his invaluable assistance in the development of my data collection tool.

Lastly, I would like to thank my family for their unconditional support and encouragement during my Ph.D. work at Georgia Tech and throughout my quest for privileged education.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

SUMMARY

This dissertation aims at developing and testing a new method that can better capture preferences for multistate health profiles. The motivation arose from the failure of the QALY (Quality-Adjusted Life Year) model in adequately capturing preferences in multistate health profiles. The current QALY-based technique captures preferences for multistate health profiles by evaluating each health state in the profile independently of other states. As the past literature showed, this additive independence condition does not hold in practice and hence such approach is inadequate. To address this issue, this study proposes a novel approach to measure preferences for multistate health profiles by looking at two consecutive health states at a time. It hypothesizes that an evaluation of the future health state is dependent or "conditioned" on the level of the preceding, or current, health state. Characteristics of the current health state that are suspected to impact the resulting “conditional preference scores” for future health state are systematically explored in a carefully designed empirical study. The interested factors include duration of the current health state, direction of change and amplitude of change between the current and future health states. A 2^3 full factorial design with three replications is used to explore main effects and their interactions. In a subsequent experiment, this study tests whether the proposed technique, which assesses "conditional preference scores" for discrete health states, can better predict preference scores for an

entire health profile than the current unconditional QALY-based technique. In this subsequent study, duration-weighted conditional preference scores, duration-weighted unconditional preference scores, and duration-weighted holistic preference scores are compared for 10 hypothetical health profiles. Visual analog scale is used as an elicitation technique throughout the experiment. The ultimate goal of this study is to enable more accurate cost-effectiveness analyses, which sequentially will lead to better healthcare resource allocation decisions.

This dissertation concludes with the results and discussions of the effects of the current health state characteristics on the preference evaluation of future health state as well as the potential of the proposed technique in capturing preferences for multistate health profiles. Implications for other related fields and future research are also discussed.

CHAPTER 1

INTRODUCTION

Today, measuring health outcomes is a crucial matter in medical decision making. When clinicians and patients make clinical decisions such as choosing medical treatments, they have to base their judgment on a quantifiable gain or loss in health. The ultimate goal of medical treatment is not to increase particular clinical parameters or to cut costs but to increase life expectancy and/or the quality of life of patients. Clinical outcomes defined in terms of mortality or physiological measures such as blood pressure or diagnostic test results are generally insufficient for formulating a judgment and/or making a decision. Patients' preferences for outcomes and risks of treatment need to be captured and explicitly included when contrasting alternative treatments for making medical decisions. At the population level, capturing and aggregating preferences is also often deemed necessary for evaluating new treatments, health services or medical technology. Lack of including such information may result in suboptimal decisions that do not conform to individual or societal preferences.

The concept and techniques of utility theory have been applied for health outcome measurement in order to incorporate patients' preferences and risk attitudes. Such utility measurement techniques have been developed and applied, to a large degree within the context of "chronic health states". A chronic health state is generally defined as a health state that stays constant over a relatively long period of time. Most real-life situations, however, challenge the assumption of a *constant* health state. Chronic diseases, even when treated, are generally not stable but lead to health status deterioration over time. Health states do not remain at the same level over lengthy periods of time even in healthy individuals.

The most widely applied model for health outcome measurement in medical decision analysis is the quality-adjusted life year (QALY) approach. The QALY model has emerged as the gold standard for health outcome measurement (Russell, Gold, Siegel, Daniels, and Weinstein, 1996). QALYs are extensively used in cost effectiveness analysis, an economic analysis technique widely used in health policy. Both life expectancy and quality of life are taken into account in a QALYs measure. The number of QALYs is typically obtained by multiplying life expectancy by a numerical weight associated with a constant health state experienced during the remaining life expectancy. The weight is a number between 0 and 1 where 0 is defined as “death” and 1 as “perfect health”. On this scale, the weight associated with a health state represents the health-related quality of life (HRQOL) of such health state. The product of the HRQOL weight and the life expectancy is a measure of the desirability of the health state experienced during the life expectancy. For example, as shown in Figure 1.1, an individual who has a life expectancy of 20 years with a disease that has a HRQOL weight of 0.7 is valued at 20 x 0.7 = 14 QALYs

Extending the approach to sequences of chronic health states such as the sequence shown in Figure 1.2, one typically calculates the desirability of such a sequence by taking the sum of all products of duration and health weight corresponding to the health states in that sequence. For example, an individual with a health profile shown in Figure 1.2 would value that sequence at [(1x8) + (0.7x5) + (0.4x7)] = 14.3 QALYs.

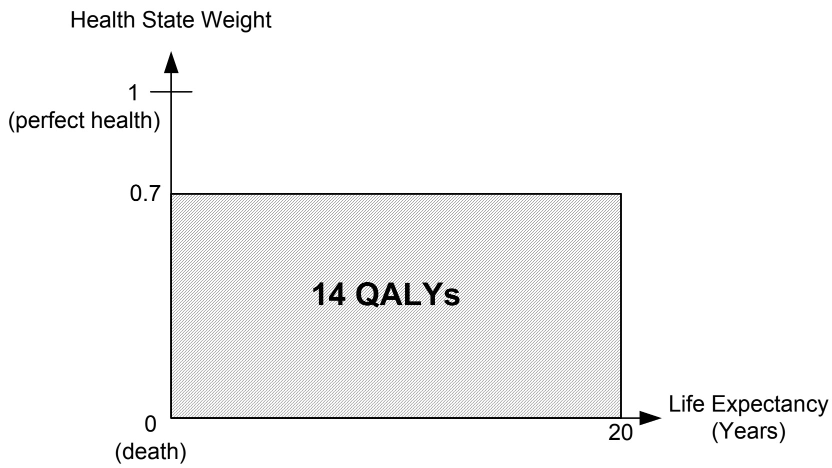


Figure 1.1: Illustration of QALYs in the case of constant health state

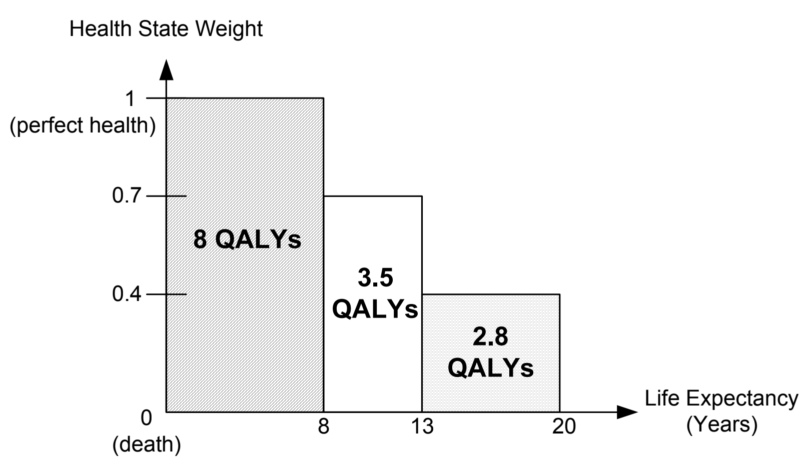


Figure 1.2: Illustration of QALYs in the case of non-constant health profile

Since the theoretical construct of QALY was developed from utility theory, the QALY model has been accepted as one of the health outcome measurements that can represent patient's preference. However, especially when health states vary, equal numbers of QALY between health profiles do not necessarily indicate that the health profiles are equally preferred. For example, consider the two profiles of health states over time depicted in Figure 1.3. Both Health Profile 1 and Health Profile 2 have been designed to produce the same amount of QALYs (the area under each curve). Yet, the two profiles are clearly different with profile 1 being very "steady" over time while profile 2 jumps up and down. Following the QALY calculations however, Health Profile 1 and Health Profile 2 would be determined to be equally preferred. However, it is plausible that the variation in health states, as shown in Health Profile 2, could have an effect on one individual's preference. The QALY framework fails to capture this.

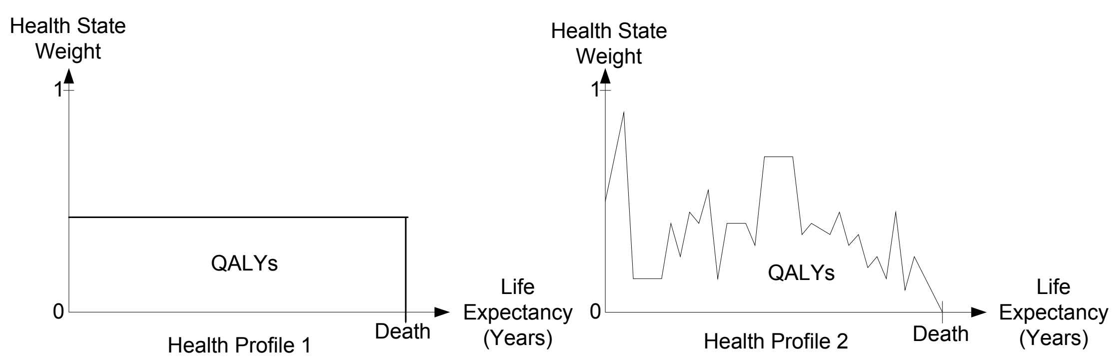


Figure 1.3: An example of potential failure of QALY in capturing preference due to variation in health states

The example above illustrates the potential effect of "steadiness" of health profile over time on patients' preference. Obviously, health profiles in reality do not stay steady over time. Thus, health profiles similar to Health Profile 1 in Figure 1.3 are unlikely to occur. Besides the steadiness of health profile, other factors defining the “pattern,” or overall shape, of the health profile might have an effect on patients’ preference. Consider another example in Figure 1.4, both health profiles are unsteady but they have different patterns. They have been designed to have equal amount of QALYs. Thus, based on an implication by QALYs, Health Profile 1 and Health Profile 2 in Figure 1.4 would be judged as equally preferred. However, this is not necessarily true. People may prefer one pattern over another even though they both produce the same amount of QALYs. There are many potential factors defining the pattern of health profiles that might affect patients’ preference. The QALY framework fails to account for these factors. Calculating QALY for health profiles with multiple health states by simply adding up QALYs corresponding to each constituent constant health state may lead to inaccurate values that misrepresent people’s preference.

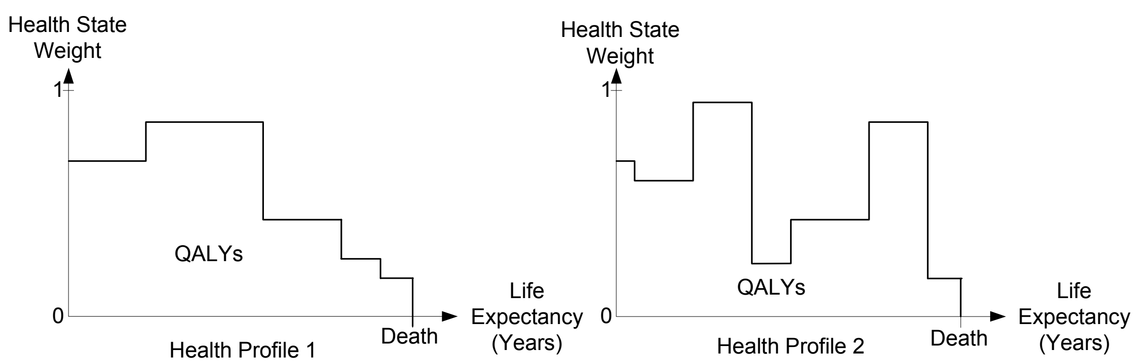


Figure 1.4: An example of potential failure of QALY in capturing preference due to the patterns of the health profiles

In order for the QALY model to represent individual's preference regarding a constant health state of interest, three conditions are required: mutual utility independence of life years and health state quality, constant proportional tradeoff property, and risk neutrality over life years (Pliskin, Shepard, and Weinstein, 1980). Moreover, to simply add QALY's representing each constant health state in a health profile comprising multiple health states, the additive utility independence requirement needs to be added to those three conditions (Bleichrodt, 1995). Additive independence requires that the preference for one health state is independent of preference for other health states in the sequence of the health profile. Thus, if additive utility independence holds, Health Profile 1 and Health Profile 2 in Figure 1.4 would be equally preferred.

However, results from several studies found a violation of the additive independence assumption in the calculation of QALYs (Richardson, Hall, and Salkeld, 1996; Kupperman, Shiboski, Feeny, Elkin, and Washington, 1997; Spencer, 2003). Hence, in the case of multistate health profiles, calculating QALYs by decomposing the health profiles into several constituent health states and adding up QALYs will most likely not be a valid technique. In this study, specific factors which are believed to influence preference assessment for multistate profiles are identified and will be empirically studied. In addition, a novel approach for combining assessments will be derived and studied to permit overall assessment of multistate profiles. Chapter 2 provides an extensive literature review of these findings and other relevant topics.

CHAPTER 2

REVIEW OF LITERATURE

2.1 QALY Framework – Theoretical Background

The quality-adjusted life year (QALY) model is a measurement technique for health outcomes that takes into accounts both quality and quantity of life. It is the arithmetic product of life expectancy and a utility-based measure of quality of life of the remaining years of life. QALY has been developed in the 1970's (Fanshel and Bush, 1970). The original theoretical properties of QALY are summarized in a paper by Pliskin et al. (1980). They show that QALY is a valid utility function, which represent individual preferences, if three conditions hold. These conditions are the followings:

1. Mutual utility independence (MUI) of life years (T) and health state (Q)

This assumption means that preferences for gambles over either one of the two attributes, with the other attribute held at a fixed level, do not depend on the particular level of that other attribute. For example, an arthritis patient does not judge his own health state differently because he has five or twenty years remaining in his life. If MUI holds, one can construct a multiattribute utility model for the health profiles (Q,T) as the followings:

$$U(Q,T) = a \cdot U(Q) + b \cdot U(T) + (1-a-b) \cdot U(Q) \cdot U(T)$$

where U(Q,T) is utility of health profile (Q,T) , U(Q) is utility of health state Q, U(T) is utility of life years T, a and b are scaling constants.

2. Constant-proportional tradeoff property

The proportion of remaining life that one would trade-off for a specified quality improvement is independent of the amount of remaining life. For instance, consider the situation where one asks an individual to trade off an amount of time of his/her remaining years of life in order to have a perfect health versus the poorer one. If he/she gives up 10 years out of 20 remaining life years, he/she would equally give up 5 out of 10, 2.5 out of 5, and so on. Thus, the proportional trade-off is constant, in this case always half of his/her remaining life years.

3. Risk neutrality over life years

This assumption means that the utility function for life years is linear. If risk-neutrality over life years holds in all health states, MUI and constant proportional trade-off will also hold (Johannesson, 1994).

The above three assumptions are the requirements for the standard QALY model which assumes risk neutrality with respect to life duration. However, the assumption of risk linearity is not empirically realistic. For example, McNeil, Weichselbaum, and Pauker (1978) found that patients with bronchogenic carcinoma had moderate risk aversion over life years. Stiggelbout, Kiebert, Kievit, Leer, Stoter, and de Haes (1994) found mild risk aversion in male patients with testicular cancer. Additionally, Verhoef, de Haan, and van Daal (1994) conducted a study with healthy women and found risk aversion over life years, but risk-seeking preferences over gambles involving short durations. On the contrary, Mehrez and Gafni (1987) found risk aversion when the length of the durations increased. Thus, the violation of risk neutrality in the standard QALY model would lead to an invalidity of QALY as a representation of individual's preference.

However, QALYs can be defined in either a risk-neutral (standard QALY model) or a more general risk-adjusted form (generalized QALY model) as developed by Pliskin et al. (1980), depending on whether the decision maker is risk neutral or not with respect to uncertainty regarding life years. If the decision maker is risk neutral, QALYs will be in the form as follows: Risk-neutral QALYs = U(Q,T) = H(Q) x T. However, by relaxing the assumption of risk neutrality, they define the more general risk-adjusted QALYs as follows: Risk-adjusted QALYs = $U(Q,T) = H(Q) \times [T]^r$

H(Q) is the quality weight, measured on a scale between 0 (death) and 1 (full health), and r is the risk parameter that defines the shape of the utility function for quantity of life. Obviously, if the subject is risk neutral, r = 1. The number of QALYs is the multiplication of the quality weight and the number of years in the health state. If mutual utility independence and constant proportional trade-off holds, then risk-adjusted QALYs as defined by Pliskin et al. (1980) are a valid utility function representing preferences over constant health states (Johanesson, 1994).

Moreover, Johanesson (1994) further classifies risk-neutral QALYs and risk-adjusted QALYs into 2 sub-types, each category being defined by the specific technique used in assessing the quality weight. They are the followings:

1. Risk-neutral QALYs

- If using the time-trade-off (TTO) method in assessing the quality weight, risk-neutral QALYs is defined by the following equation: Risk-neutral QALYs = U(Q,T) = V(Q) x T, where V(Q) is the value function of health state Q.
- If the standard gamble techniques are used in assessing the quality weight, risk-neutral QALYs is defined by the following equation: Risk-neutral QALYs = U(Q,T) = U(Q) x T

If risk neutrality over quality of life holds, V(Q) is equal to U(Q) and both the quality weight and the number of QALYs will be the same for both the standard gamble method and the time-trade-off method.

2. Risk-adjusted QALYs

By relaxing the assumption of risk neutrality over life years, risk-adjusted QALYs can be defined in the QALY framework as follows: U(Q,T) = U(Q)·U(T). Johanesson (1994) classifies risk-adjusted QALYs into 2 types regarding the techniques used in assessing the quality weight as follows:

- If using the TTO method in assessing the quality weight, risk-adjusted QALYs is defined by the following equation: Risk-adjusted QALYs = U(Q,T) = $[V(Q)\cdot T]^{r}$
- If the standard gamble techniques are used in assessing the quality weight, risk-adjusted QALYs is defined by the following equation: Risk-adjusted QALYs: $U(Q,T) = U(Q) \cdot T^{r}$, where r is the risk parameter. If the subject is risk neutral, r = 1.

More specifically, U(Q) is equal to $V(Q)^{r}$ and V(Q) is equal to $U(Q)^{1/r}$, if the constant-proportional tradeoff property holds (Johannesson, 1994).

Besides challenging the risk neutrality assumption, several studies on the validity of the other assumptions have been preformed. For example, Miyamoto and Eraker (1988) tested the mutual utility independence assumption and found empirical support for this assumption. Bleichrodt and Johannesson (1996) performed empirical tests on utility independence and constant proportional tradeoff. They found that without adjustment for imprecision of preference (imprecision adjustment was suggested because of the unfamiliarity of the subjects regarding both the health states and elicitation methods), 22.8% of the subjects satisfied constant proportional tradeoff, 13.4% satisfied utility independence, and 5.8% satisfied both assumptions. However, with the imprecision adjustment, 90.1%, 75.8% and 88.8% of the subjects satisfied constant proportional tradeoff only, utility independence only, and both assumptions, respectively. The authors concluded that the constant proportional tradeoff holds roughly and utility independence holds, but in a much weaker way. Pliskin et al. (1980) reported 25 pairs of time-tradeoff responses from 10 subjects in hypothetical questions concerning the relief of different levels of angina pain. They found that only four out of 25 pairs were consistent with the constant proportional tradeoff assumption.

2.1.1 Zero-Condition

Instead of the three assumptions established by Pliskin et al. (1980), which require knowledge of concepts from utility theory, Bleichrodt, Wakker, and Johannesson (1997) suggested a more elementary and fundamental characterization of QALYs that can relax Pliskin et al. (1980)'s assumptions. They found that risk neutrality together with the "zero-condition" are sufficient to imply the existence and validity of the QALY model.

The “zero-condition” indicates that all health state levels are equivalent, from a quality of life perspective, for a zero duration of life years. Zero condition is a condition that seems unavoidable in the medical context. Thus, the only assumption that is needed to imply QALYs is the risk neutrality for all health states.

However, there is ample empirical evidence showing a violation of risk neutrality as previously described. A generalized QALY model that can relax the assumption of risk neutrality was established to solve risk neutrality issue. A generalized QALY model has the following form: U(Q,T) = V(Q)W(T), where U(Q,T) is the utility of the health profile (Q,T), V(Q) is the value function of health state Q, and W(T) is the function that values life duration and can be nonlinear, with W(0) = 0. Instead of the risk neutrality condition, Miyamoto, Wakker, Bleichrodt, and Peters (1998) suggested another condition, “standard gamble invariance (SG invariance)”, in order to relax the risk neutrality condition. SG invariance basically says that, if Q and Q’ are unequal to death and *p* is the probability equivalent of (Q, T) with respect to (Q, Y) and (Q, Z), then *p* is also the probability equivalent of (Q’,T) with respect to (Q’, Y) and (Q’, Z) (Miyamoto et al., 1998). Without risk neutrality, a generalized QALY model holds if and only if the zero-condition and SG invariance hold.

2.1.2 Violation of Expected Utility Theory Assumption when Death is Considered as an Outcome

Expected utility theory or the von Neumann-Morgenstern expected utility theory is the foundation for health outcome assessment and measurement. A utility function actually exists when axioms of rationality hold for the preference relations being studied.

Three axioms of expected utility theory (von Neumann and Morgenstern, 1947) which have been known to be normatively compelling rules for rational decisions under uncertainty are as follows (*Note: "$\succ$" denotes an individual's "is preferred to" relation, and "~" denotes the indifference relation, X is the set of outcomes, and Δ (X) is the set of probability distributions over X)*:

1. Weak order

$\succ$ is asymmetric ($p \succ q$ => not [$q \succ p$]) and both $\succ$ and ~ are transitive (if $p \succ$ (or ~) q and $q \succ$ (or ~) r then $p \succ$ (or ~) r for all p, q, r $\in \Delta(X)$.

2. Independence

For all p, q, r $\in \Delta(X)$ and any $\alpha \in]0, 1[$, then $p \succ q$ if and only if $\alpha p + (1-\alpha)r \succ \alpha q + (1-\alpha)r$.

3. Continuity Axiom

If p, q, r $\in \Delta(X)$ such that $p \succ q \succ r$, then there exist α and $\beta \in]0, 1[$ such that $\alpha p + (1-\alpha)r \succ q \succ \beta p + (1-\beta)r$.

Thus, in addition to the three required assumptions of QALYs described previously, how adequately QALYs represent preference over health states also depends on whether they are consistent with von Neumann and Morgenstern's expected utility theory. If the axioms of von Neumann and Morgenstern's expected utility theory hold true, decision makers should be able to make decisions that are consistent with their underlying preferences for each possible health outcome. However, in medical decision making, when death is considered as an outcome, a violation of the continuity axiom may arise. For example, consider three health outcomes: perfect health, having a cold, and death. Perfect health is preferred to having a cold, which is preferred to death. In order to satisfy the continuity axiom, one should be able to find α and β (between 0 and 1) for the gamble in Figure 2.1. However, for most individuals, death would be considered as an infinitely undesirable outcome, and it would seem impossible to find α strictly greater than 0 such that an individual would be willing to risk death in order to avoid the state of having a cold. Therefore, this would be a violation of the continuity axiom.

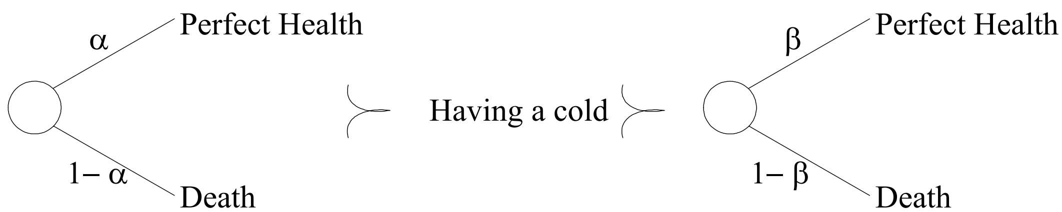


Figure 2.1: Illustration for a possible violation of the continuity axiom when death is part of the outcome set

In the case of multistate health profiles, QALYs are calculated as the sum of all products of time and health preference weight for all health states representing the health profile. Bleichrodt (1995) has shown that in order to obtain a valid measure of QALY under expected utility theory for multistate health profiles by adding up QALYs calculated from constituent states, it is necessary that the assumption of additive independence holds. As mentioned above, additive independence requires that the preference for one health state is independent of preference for other health states in the sequence of the health profile (Bleichrodt, 1995).

Regarding the additive independence assumption, there have been several empirical research studies that explored the validity of the additive independence assumption. The review of these studies is described in the next section.

2.2 Testing the Additive Independence Assumption

Richardson et al. (1996) examined the validity of the additive QALY model in a 16-year post-mastectomy health profile represented by a gradual deterioration composed of three health states: moderate side effects during the first five years, mild side effects for the next 10 years, but then breast cancer would recur and the patient would experience severe side effects during the last one year. Sixty-three female respondents participated in the study. Rating scale, time-tradeoff and standard gamble techniques were used to assess utility for each health state and the holistic utilities for the health profiles. Preference scores from constituent states were combined to estimate scores for the health profile using a discount rate of 3% and 9%. They found that holistic preferences for the multiphase health profile (whether assessed with a rating scale, time-tradeoff, or standard gamble) were significantly different from composite preferences derived from the constituent health states, irrespective of the discount rate applied.

Kupperman et al. (1997) also investigated whether preferences for multiphase health states can be approximated by preferences from constituent health states. A hundred and twenty-one female subjects were asked to assess their preferences on eight health profiles, each composed of three to four health states, in the context of prenatal diagnosis choices: chorionic villus sampling and amniocentesis, by using visual analog scaling and standard gamble techniques. They explored if a different statistical formulation could be derived to predict preference scores for health profiles from their constituent health states preference scores. They found that a duration-weighted additive model, in which the preference scores are duration-weighted and summing up, as the conventional QALY model does, was not predictive. A multiple regression model that derived from statistically inferred weights could predict the preferences for the profiles better than the duration-weighted model.

MacKeigan, O'Brien, and Oh (1999) used the time-tradeoff technique to compare preference scores for the same lifetime paths between holistic and composite assessment. A hundred and one participants with type 2 diabetes assessed their preferences regarding four hyperglycemic treatment profiles lasting 30 years, composed of 8 discrete treatment states. They failed to find any significant differences between holistic and composite scores, which conflicted with the results from the studies by Richardson et al. (1996) and Kupperman et al. (1997). However, the health profiles used in MacKeigan et al. (1999) were different in that they consisted of progressive minor deteriorations in states while the health profiles in the other studies consisted of critically different health states. Moreover, the authors concluded that another reason would be because the profiles in the study were too similar. They recommended that future research should be repeated with profiles that are more distinct and where sequencing effects are likely to be larger.

In Spencer (2003)'s study, three health states defined with the EuroQol classification system were used in each health profile. Each health profile in the study had 10 years duration and composed of up to three different health states with the duration of three years, three years, and four years. Two tests were conducted: test of additive independence and test of the additive model. Twenty-nine subjects participated in the study. The violation of additive independence was found in the first test. However, in the additive model test, only one of the two versions resulted in a rejection of the additive model. Thus, Spencer could not conclusively reject the additive model. The author suggested that a larger sample size might allow the test to be able to detect significant differences in the results. Also, comparisons of utilities based on holistic elicitation procedures and constituent states elicitation were performed. The results showed that two out of the seven profiles exhibited a significant difference between holistic and constituent states elicitation, which implied that additive independence assumption was violated.

The studies previously described clearly show the violation of the additive independence assumption. When there are multiple outcomes in the profile, people's preference scores of the outcome profile cannot be evaluated by simply adding scores assessed from segmented outcome assessments. There are a number of studies that explored influential factors on people's preferences, which could lead to the violation of additive independence assumption. The review of the studies for each potential factor is described below.

2.3 Potential Influential Factors

2.3.1 Rate of Change

Hsee and Abelson (1991) performed experiments to find a relationship between satisfaction and rate of change of the outcomes or what they called "velocity". Contexts of the outcome that were used in the study are gambling (probability that subjects would win the game), class rank (the percentile standing in a hypothetical class), and stock (a stock price that the subjects were hypothesized to invest into). They found that satisfaction is positively related to not only the actual outcome position but also the rate of change (or velocity) of the outcomes over time. The more positive (negative) the rate of change is, the greater (less) the satisfaction will be. This implies that for an improving sequence of outcomes, a more rapidly improving rate is preferred. On the other hand, for a deteriorated sequence of outcomes, slower deteriorated pace (less negative rate of change) is preferred. For example, in the context of health, people would look forward to recovering rapidly from having a poor health state after receiving a treatment. Moreover, when they have a sickness, they would not desire to have a disease that worsens their health state hastily. Those results are certainly sensible in a medical context.

Additionally, Hsee, Abelson, and Salovey (1991) investigated factors that influence the relative weighting of position and velocity in satisfaction. They found that the framing of the outcome, the perceived purpose of the behavior associated with the outcome (motive), and the control of the outcome had an effect on the relative importance of the actual value (position) and the change in the value (velocity). The relative importance of the velocity was larger than that of the position when the outcome was framed in terms of change rather than in terms of an overall position, and vice versa. Moreover, if the activity is intrinsically motivated (e.g. performed because it is enjoyable), the satisfaction will be more influenced by the change in the outcomes. On the other hand, if the activity is extrinsically motivated (e.g. performed to achieve some external outcomes), people are more concerned with the final outcome of the activity. Furthermore, if the outcome is perceived as internally controlled, the relative importance of velocity will be larger since it indicates the performance of the person who performs the activity. In contrast, if the outcome is perceived as out of control, the final outcome will be more important.

Chapman (1996a) also found a significant effect of rate of change or “slope” on the preference in both health and money domains. Forty undergraduate students evaluated 16 different patterns of sequences of health and money by using a 0 to 100 scale. Five of the 16 sequences showed health or money decreasing at constant rates that varied over the sequences. Another five sequences showed health or money increasing at constant rates, with the different rates over the five sequences. Thus, these 10 sequences had five different slopes or rate of change for each direction of change over time (increasing or decreasing). One of the 16 sequences had constant money or health over time. Another one had money or health at a maximum for the first half duration and at a minimum for the last half duration. Another four sequences represented more realistic sequences. Two of them decreased gradually at the early part of duration and more steeply at the later part. Another two sequences increased steadily until two-third of the duration and then dropped to a low and constant level. Rating scores from the 10 sequences that had constant rates of change varied over the sequences (five are increasing sequences and another five are decreasing sequences) were analyzed to study the main effect of slope. It was found that slope or rate of change is one of the significant factors that impact their rating scores. It indicated that subjects preferred gradually increasing or decreasing sequences to those with steep slopes. These results were conflicted with the findings by Hsee and Abelson (1991), which suggested that the more positive (negative) the rate of change is, the greater (less) the satisfaction will be. Or in other words, for an increasing sequence, the steeper the slope is, the greater the satisfaction will be. For a decreasing sequence, the gradually decreasing sequence is preferred to the decreasing sequence with steep slope. However, these two studies were different in that the study by Hsee and Abelson (1991) did not controlled for the total number of units of outcome over a specific period of time while the study by Chapman (1996a) did. Thus, preference for higher rate of change in positive outcome in the findings by Hsee and Abelson might be a result of higher amount of the outcomes received within the specific period of time.

Ariely (1998) also found a significant effect of rate of change in a study of retrospective pain evaluation. Sixty-four different patterns of heat stimuli with two different slopes (gradual and steep) varied among the patterns were experienced by 20 subjects through the probe of Thermode (a device that controlled level of heat) on their forearms. The subjects gave an overall pain evaluation of each pattern at the end of experience with each stimulus. The results showed that the intensity slope or rate of change of the final part of the experience had positive impact on the evaluations. The subjects rated the experience as having higher pain when the intensity steeply increased than when it gradually increased.

2.3.2 Peak, Final Outcome, and Duration of the Profiles

When making global evaluation of the experience, the most logically compelling rule is a rule of temporal monotonicity: adding aversive experience to the episode worsens the global evaluation and adding pleasant experience enhances the evaluation (Kahneman, Fredrickson, Schreibner, and Redelmeier, 1993). However, there were several studies showing the violation of this rule. Adding aversive moment to the episode was found to increase the evaluation positively if the moment added is less aversive than the previous moment even though it makes the total duration of discomfort longer (Varey and Kahneman, 1992; Kahneman et al., 1993, Fredrickson and Kahneman, 1993; Redelmeier and Kahneman, 1996; Baumgartner, Sujan, and Padgett, 1997; Langer, Sarin, and Weber, 2000; Ariely and Loewenstein, 2000; Diener, Wirtz, and Oishi, 2001; Schreiber and Kahneman, 2000). A number of empirical works have demonstrated that retrospective pain evaluation is influenced essentially by the peak and final moment of the experience and not significantly impacted by the duration of the painful experience itself. In medical decision making, retrospective pain evaluation is an important matter since it reflects patient's memories of how painful the treatment was and could impact their decisions regarding future treatments.

Varey and Kahneman (1992) examined intuitions relating to outcomes extended over time and found that extended aversive experiences were not evaluated with the utility integration rule (the duration and the trend of the experience should jointly determine the global evaluation). In one experiment, when the duration of experience was emphasized, intuitions about the global evaluation of extended aversive experiences were influenced by duration rather than the intensity of the experience. Forty-eight subjects were asked to evaluate series of unpleasant experiences (carrying a suitcase for 200, 550 or 990 yards; standing within a knee-high metal ring 18-inch-diameter for 5, 9, or 13 minutes; wearing a helmet and breathing apparatus for 5, 9 or 13 minutes; reading while exposed to a sound for 20, 35, or 50 minutes). Duration was found to have an effect on the disutility evaluation as a concave function, even for experiences that intensified over time. In another experiment, when the trend was made explicit, the global evaluations were primarily influenced by the trend of the experience and were insensitive to duration.

In this experiment, 46 subjects were asked to assess their own global evaluation of another person's experience of different discomfort rating series using a 0 to10 scale. The series varied in the intensity and trend and ranged from 15 to 35 minutes long. It was found that discomfort rating series that had better ending were judged much less aversive than those with worse ending moment even though those series were shorter in total duration. The authors hypothesized that in the evaluation of extended experiences, the subjects simplified the task by focusing on some dimension(s) as representative of the whole experience. The dimension(s) that was(were) selected depended on how the information was presented. Another finding was that when the trend of experience influenced the global evaluation, the peak discomfort had a large impact on the evaluation while the duration was neglected. This implies that adding pain to the end of the sequence can lower the evaluation if the added duration has lower pain than the preceding level. This finding was consistent with the finding by Ross and Simonson (1991), which stated that the preference of the decision makers was affected by the final moment of the sequence.

In a study by Kahneman et al. (1993), 32 subjects were asked to experience two trials: one is to immerse one hand in 14°C water for 60 seconds and the other one is to immerse the other hand at 14°C for 60 seconds and then the temperature was gradually increased to 15°C in another 30 seconds (total duration was 90 seconds). Sixty-nine percent of the subjects when asked to repeat one of the two trials chose to repeat the longer one. Moreover, most subjects indicated that the long trial had less overall discomfort, was less cold at its extreme moment, and was less tough to cope with. The correlation of 0.16 between trial durations and choice selected confirmed the neglect of duration. Furthermore, results from Fredrickson and Kahneman (1993)'s study confirmed the influence of peak and final moment and the negligence of duration on on-line and retrospective evaluation. Thirty-two subjects viewed a total of 16 film clips, varying in content from pleasant to aversive, ranging in length from 29 to 125 seconds. Subjects provided on-line rating during watching each clip and retrospective rating after watching each clip. The results were similar to the previous studies that the global evaluations of were impacted by the peak and final moment of the episodes and very little by the duration. Similar results were also found in Redelmeier and Kahneman (1996)'s study. They assessed on-line and retrospective experience evaluations from 154 patients who underwent colonoscopy and 133 patients who underwent lithotripsy and found that the patient's memories of painful medical treatments were mainly affected by the peak and final intensity of the procedures. The duration of the treatments did not significantly impact the retrospective evaluation.

Ariely and Zauberman (2000)'s study also provided evidence for a significance of trend and a preference for happy ending. Fifty-four students were exposed to different intensity patterns of annoying sounds and asked to give the retrospective rating on the overall annoyance. It was found that subjects rated patterns that had final increasing trends as more annoying than those that had final decreasing trends. Moreover, this study investigated the effect of segmentation on the evaluation of the experience. They found that when the experience was segmented into discrete parts such that it will be more like multiple experiences, the evaluation would be based more heavily on the intensity of the parts themselves and less on the pattern. Additionally on-line measurement was found to reduce the impact of the pattern on the overall evaluation as was also found by Ariely

(1998). Ariely and Zauberman (2000) suggested that subjects segmented their experience while they were doing on-line rating. Thus, both on-line measurement and segmentation of the experience can reduce the impact of the pattern on the overall evaluation.

Ariely and Carmon (2000) performed a study on patients with bone marrow transplant. The patients were asked to rate the level of pain on a scale of 0 (no pain) to 100 (the worst pain) on every hour from 8am to 6pm on the experiment day. Also, at the end of the day, they were asked to give an overall evaluation of the pain on that day. It was found that based on the hourly rating that they reported, the final pain intensity and the slope were the significant predictors for the overall evaluation. However, in contrast to other studies (Varey and Kahneman, 1992; Fredrickson and Kahneman, 1993; Redelmeier and Kahneman, 1996; Baumgartner et al., 1997; Schreiber and Kahneman, 2000; Langer et al., 2002), peak intensity was not found to be significant. The author explained the conflict by advancing that since the subjects were long-term patients in the hospital, the peak intensity that they had on the experiment day may not be clearly differentiated from the peak intensity they had on the days prior to the experiment. The small effect of peak intensity was also found in the study by Ariely (1998) where the peak intensity was found to have a significant effect but much smaller than the slope and the final intensity. The author explained that the subjects were exposed to too many trials (a total of 110 trials in two experiments) and hence, the salience of the peak intensity, or lack thereof, may have moderated its impact on the overall evaluation.

2.3.3 Sequence of Outcomes

Sequence effect denotes a change in preference that is influenced by different patterns of outcomes. This includes the difference of the preference when assessing health outcome that is presented as an individual outcome versus a sequence of outcomes. One of the factors that play an important role in sequence effect is time. People may have different preferences regarding the same outcome for different points in time. A number of research studies have shown that people, when faced with single outcome choices, exhibit positive time preference (Benzion, Rapport and Yagil, 1989; Chapman 1996a, Chapman and Elstein, 1995; Thaler, 1981). They prefer to have positive outcomes sooner than later, and prefer to delay negative outcomes to later.

Lipscomb (1989) investigated time preferences over 96 different hypothetical health profiles. Each health profile began at age 25, started with excellent health, followed by one of two poor health states (either condition X or Z) with four different durations (one month, one year, three years, and five years), starting after one of six different delays (no delay, three years, five years, seven years, 15 years, and 25 years). Following condition X or Z would come either excellent health until age 75 or immediate death. Fifty-two undergraduate students were asked to assess their preferences for these health scenarios by using categorical rating as well as standard gamble techniques. Regression analysis was used to explain these scores with delay of onset, health condition, duration, exit state, respondent characteristics (e.g. own health status, smoking behavior), and assessment technique. The results showed that the impact of the state of poor health on the evaluation of the health scenario was reduced as the poor health was delayed further into the future, implying a positive time preference. Preference scores increased as the length of delay increased.

However, when the outcomes are presented in sequence, people tend to display negative time preference in which improving sequences of outcomes are preferred to declining sequences (for example, Loewenstein and Sicherman, 1991; Loewenstein and Prelec, 1991, 1993; Varey and Kahneman, 1992; Ross and Simonson, 1991).

A study by Loewenstein and Sicherman (1991) showed that a majority of the subjects exhibited negative time preference in the context of wage profiles over a 5-year period. They preferred to have increasing wage profiles rather than a declining or a constant one even though the net present value of improving wage profiles were smaller than that of declining or constant wage profiles. Moreover, Loewenstein and Prelec (1991) performed experiments on choices of restaurant and time preference. A majority of the subjects who reported as having preference for a French restaurant over a Greek restaurant preferred to have dinner at French restaurant sooner rather than later when the choices were presented as a single outcome. However, when the outcomes were presented in sequence (e.g. dinner at a Greek restaurant followed with dinner at a French restaurant, on different days of course), more than half of the subjects preferred to have dinner at a Greek restaurant first and at a French restaurant later. This also shows an empirical evidence of negative time preference for the sequence of outcomes. Furthermore, a study by Loewenstein & Prelec (1993) found that a majority of the subjects exhibited negative time preference when the decision frame emphasizes the sequential nature of the outcomes. However, when single outcomes are emphasized, positive time preference was displayed instead.

Chapman (1996a) explored preferences for both sequences of health and monetary outcomes. Forty students were asked to give preference scores (score 1 (least preferred) to 10 (most preferred)) for 32 sequences: 16 in the health domain and 16 in the money domain. The sequence length for the health domain was 60 years, and 5 years for the money domain. It was found that decreasing sequences of health were preferred to increasing sequences. However, the preferences for the decreasing sequences were weaker in the money domain. In the money domain, increasing sequences were not found to be preferred to decreasing sequences, which was inconsistent with the result from Loewenstein and Sicherman (1991)'s study. Chapman explained that Loewenstein and Sicherman (1991) used relatively short sequences (5 years) whereas a lifetime sequence was used in her study. As shown in another experiment by Chapman (1996a), one year and lifetime sequences were used in both health and money domains. Subjects reported to have preferences for improving sequences over the decreasing ones for both health and money in the one-year sequence whereas in the lifetime sequence, the decreasing sequences were preferred in health and the improving sequences were preferred in money. The results indicated that preferences for sequences are influenced not only by domain but also by the sequence length. Chapman suggested that the preferences for improvement exhibited were driven by their expectations. When considering a long time horizon such as lifetime, they expect a perfect health early and gradually declining health as they are getting older. Also, they expect an income that increases over their lifetime. The subjects used their expectation as a reference point and judged their preference by considering how close the profile in question was to their reference point.

Later, Chapman (2000) further investigated the idea that expectations drive preferences in three experiments. A hundred subjects in the first experiment were asked to make choices between improving and declining sequences of six questions of headache pain. The six questions were differed in sequences lengths: one hour, one day, one month, one year, five years, and 20 years. Ninety subjects in the second experiment were asked three choice questions of athletic ability. Each question asked the subjects to make selections between improving and declining sequences. The three questions were differed in sequence lengths: one week, one year, and 20 years. A hundred and fifteen subjects in the last experiment were asked the choice questions between improving and declining sequences of three questions about facial acne and three questions about facial wrinkles. The questions differed in length of sequence: one week, one year, and 20 years. Another group of subjects gave expectation ratings for each of the 15 sequence pairs (six pairs for headache pain, three pairs for athletic ability, three pairs for facial acne, and three pairs for facial wrinkle). It was found that in each experiment, preferences generally followed the expectation pattern. The expectation ratings were significantly correlated with the percentage of subjects who preferred improving sequences. For the headache pain, the percentage of the subjects who chose the improving sequence slightly increased with longer sequences. However, the expectation rating across the sequence lengths did not follow this pattern. In the athletic ability and the facial complexion, the percentage decreased with longer sequences. The author concluded that expectation influences preferences for some specific health outcomes, rather than by decision domains (e.g. health versus money).

Krabbe and Bonsel (1998)'s study also showed evidence for a sequence effect in health. A hundred and four subjects evaluated 13 different EuroQol-defined health states using each of two variations of the TTO technique (ordinary TTO and reverse TTO). With the ordinary TTO, assessments were made starting with the "best imaginable state" first and ending with the "worse imaginable state" last. With the reverse TTO, the reverse order was used. The values for "best imaginable state" from both TTOs should be the same for each health state if the sequence effect did not exist. However, after applying a 5% discount rate, they found that values for the "best imaginable state" in the reverse TTO questions were lower than those in the ordinary TTO. Thus, this study provided additional evidence for a sequence effect in health state measurement.

Possible explanations for preference for improving sequence (negative time preference)

Loewenstein and Prelec (1993) suggested three reasons that explained preference for improvement or negative time preference of the sequence. First, the belief of savoring and dread (Loewenstein, 1987), people prefer to have negative outcome sooner to eliminate dread which allow them to savor the best outcome until the end of the sequence. The second one was adaptation (Helson, 1964) and loss aversion (Kahneman and Tversky, 1979). People tend to adapt to changes in situation over time and they try to avoid confronting loss. Improving sequence allows them to positively adapt and always face gains when they evaluate the new outcome by comparing to the existing one. A decreasing sequence, on the other hand, implies a sequence of relative losses. The third reason was the recency effect (Miller and Campbell, 1959), when a person doing retrospective evaluation, the last moment tends to be the easiest one to be retrieved from the memory. Thus, when asking people to evaluate their preference with respect to a series of outcomes, improving sequences which put the best outcome last leads people to have a more favorable perception of the sequence as was evidenced by Varey and Kahneman (1992)'s study. Subjects showed preferences for sequences that had a happy ending. Besides, the recency effect also appeared in Ross and Simonson (1991)'s study. They investigated the effect of chronological order of positive and negative events on people's overall evaluations of their experiences. It was found that in the given occurrence of two events, the subjects demonstrated a preference for happy endings, having the positive event occurred last.

Loewenstein and Sicherman (1991) suggested that people make choices by using a rule of reference point. They compare the outcome they have with their reference point and make judgments accordingly. A preference for an improving sequence can be explained by the fact that people interpret an improving sequence as a series of gains and a declining sequence as a series of losses. They might use the previous outcome as a reference point and use it to assess and evaluate the current outcome. Chapman (1996a) also pointed out that expectation is another candidate reference point that people might use. Thus, people tend to prefer the sequence that is closer to their expectations.

2.3.4 Spreading of Outcomes

Loewenstein and Prelec (1993) found that decision makers prefer outcomes that spread across the time interval. In their study, 84% of the subjects preferred to have dinner at a Fancy restaurant on the second weekend in the total spread of three weekends, while the other two weekends would be dinner at home. However, when they were offered a fancy lobster dinner on the third weekend, 54% of the subjects preferred to have fancy dinner at French restaurant on the first weekend rather than the second weekend. In their next experiment, the time interval was then expanded to five weekends instead of three weekends. When a dinner at a French restaurant on one weekend was offered among dinners at home on the rest of the weekends, 88% of the subjects preferred to have the dinner at a French restaurant on the third weekend rather than the first. However, when a fancy lobster dinner was put at the fifth weekend, 51% of the subjects chose to have dinner at French restaurant on the first weekend rather than on the third. The results from both experiments showed that subjects tended to distribute outcomes across the time interval in decision.

Chapman (1998) performed a study of preferences to examine the effect of spreading with scenarios including both gain and loss in the context of monetary outcomes (win a prize or pay a fine), dinner (pleasant dinner or unpleasant dinner), and health-related events (a painful trip to the dentist or a pain-relieving trip to the chiropractor) in a four-week interval. In all scenarios, the majority of the subjects (range from 70% to 92%) preferred to have the event on the first and the third weekends rather than on the first and the second weekends. About 65% to 91% of the subjects preferred to have the event happen on the second and the fourth weekends rather on the third and the fourth weekends. The results were very similar to the results from Loewenstein and Prelec (1993)'s study, which supported the preferences for spreading outcomes.

2.3.5 Study of More than One Factor

Ariely (1998) explored the joint effect of factors that influenced and determined global pain evaluation. The factors that were included in the experiment were physical intensity at the end and at the peak of the experience, rate of change, the trend, and the duration of the experience. Twenty subjects were experienced 64 different heat stimuli that passed through contact of a Thermode (a device that controlled level of heat) on the inner part of their forearms. At the end of each trial, the subjects gave evaluation of the experience by using a visual analog scale. It was found that the retrospective pain evaluation was primarily influenced by the pain intensity at the end and the rate of change during the latter half of the experience. Both of the factors positively impacted the global evaluation. Longer duration of the experience increased the evaluation only when the intensity of the pain during the extended episode changed over time, but not when the intensity was constant.

2.4 Time Discounting of Health Outcomes

Another matter that plays a substantial role in medical decision making is timing of the events. Intertemporal choice is a decision that involves outcomes that occur at different points in time. People value the same event differently when it happens at different timing. However, in the conventional QALY approach, each year in the lifetime is weighed equally. Thus, if the health state stays constant over time, then the valuation of health state based on the QALY approach will be constant over time, which is not realistic. In cost-effectiveness analysis, discount rates are typically applied in order to deal with this time preference issue.

Intertemporal choice usually involves trade-offs between current and future consumption. For example, a decision to stop smoking cigarette involves trade-offs between immediate gratification of cigarette smoking and a healthier life in the future. The study of time discounting of health outcomes also allows understanding people's preventive health behaviors, which involves a choice between a (small) enjoyable immediate outcome (for example, eating a high-calorie dessert now) and a (larger) delayed outcome (for example, being healthier at an older age). People who choose a small immediate reward over a larger delayed outcome are thought of as having a large time discount rate. They value the larger outcome in the future to be smaller than the smaller outcome now. Thus, time discounting might be a possible explanation for people who fail to take preventive health behaviors.

An additional rationale for time discounting is that immediate outcomes can be expected to lead to additional positive outcomes in the near future. For example, as soon as the patient gets a treatment that will bring her from an ill state to full health, she can start enjoy life, work and earn some money sooner. Another rationale for discounting is that risks also compound over time. For example, the patient may want to have the treatment as soon as possible since there is a risk that her health may get worse or a risk of death might arise while waiting for the treatment.

Discounted utility theory, a theory by Paul Samuelson (1937), is a normative model designed to accommodate intertemporal choices. Basically, the discounted utility theory says that all of the time effect in intertemporal choice can be condensed into a single discount rate and this discount rate is constant over time. Moreover, when the outcomes are exchangeable, all domains should use the same discount rates. However, the descriptive validity of the discounted utility theory is still being debated. Research on time discounting has focused almost entirely on the validity of the discounted utility model by challenging the two main assumptions: constant discount rate over time and single discount rate for all domains. Several studies found evidence of domain independence which implies a violation of single discount rate for all domains assumption (for example, Chapman and Elstein, 1995; Chapman, 1996a, 1996b, 2002; Ganiats, Carson, Hamm, Cantor, Sumner, Spann, Hagen and Miller, 2000; Redelmeier and Heller, 1993; Cairns, 1992; Chapman, Nelson, and Heir, 1999). Moreover, a violation of the constant discount rate has been evidenced by several studies (for example, Christensen-Szalanski, 1984; Cropper, Aydede, and Portney, 1992; Dolan and Gudex, 1995).

Time preference studies have focused on various domains such as monetary outcomes or health outcomes. However, for the scope of this dissertation, this review focuses on health outcomes time preferences.

2.4.1 Health Discounting Behavior

<u>Variable discount rates</u>

Time discount rates in health outcomes have a high degree of variability. Some studies found discount rates to be very high (for example, Ganiats et al., 2000 (up to 116%); Chapman and Elstein, 1995 (up to 263%); Chapman, 1996b (up to 300%); Chapman et al., 1999 (up to 19,000%)). On the other hand, some studies found zero discount rates. For example, in Cairns (1992)'s study, 29 students were asked to self-assess the number of days in delayed severe depression that they would be indifferent between two scenarios: one had severe depression immediately for 90 days, and the other one delayed the severe depression to later (2 years, 20 years, or 30 years). It was found that overall, the time preference rates for health in this study were zero or small but positive, in the range of 0.4% to 3%. Redelmeier and Heller (1993), Dolan and Gudex (1995) also found zero discount rate in their study.

Moreover, time preferences for health outcomes were found to vary among disease conditions as shown by Ganiats et al. (2000). They studied time preferences with five different disease conditions: chicken pox, Parkinson's disease, tropical disease, migraine headache, and sterilization. Chicken pox and tropical disease were framed as losses, e.g. expose the child to chicken pox now or later, having a poor health due to tropical disease for three months now or later (six months, one year, two years, five years, 10 years, or 20 years). The migraine headache and Parkinson's disease were framed as gains, e.g. taking medication to relief now or later. While the sterilization context involved the time-risk tradeoff between two procedures. One was safe in the short run but presented a risk for cancer in the next 20 years, another one was risky in short run and did not present long-term cancer risk. Negative time preference was found in the chicken pox context, while positive time preference was found in the Parkinson's disease. However, some duration in the tropical disease showed positive time preference and some showed negative time preference. The responses to the migraine headache context showed a large discount rate (116%) while a small discount rate (3%) was found in the sterilization context.

Delay effect

Numerous researches on time discounting found that people tend to discount outcomes that occur in short delays higher than those that occur in long delays. The existence of delay effect provides support for the invalidation of the constant rate discounted utility model. Bleichrodt and Johannesson (2001) studied time preference of health outcomes with the condition of back pain. A hundred and seventy-two subjects assessed two scenarios of sequences of outcomes at a time and chose the scenario that was most preferred. The scenarios were different in delay and the duration of back pain. The results showed that time discounting rates decreased as the delays were getting longer. Another evidence for the delay effect was found in a study by Redelmeier and Heller (1993), 121 subjects were asked to assess their time preferences on three hypothetical health states: colostomy, blindness, and depression that will occur in the future (one day, six months, one year, five years, and 10 years) using standard gamble and categorical scaling techniques. They found that discount rates for all three health states were smaller for time intervals that were relatively more distant (e.g. 5 years to 10 years in this study), conflicting with the constant discount rate assumption in discounting utility theory.

Chapman and Elstein (1995) also found evidence of delay effect in both of their experiments. In the experiments, time discounting for health and money domains were studied. In the first experiment, the lengths of delay used were six months, one year, two years, and four years, while in the second experiment, one year, three years, six years, and 12 years were used. Subjects were asked to find the magnitude of delayed outcomes that would make the two options equally attractive. The results from both experiments showed that the discount rates for both health and money domains decreased as the length of delay increased. Moreover, Chapman (1996b) performed three additional experiments on time discounting in health and money. All of the results from the three experiments suggested delay effects for both health and money domains. The discount rates decreased as the length of the delay increased. In her first experiment, gains in health and money with delays of one year, three years, six years, and 12 years were employed. While in the second and third experiments, both gains and losses in health and money domains were used. The delay effects existed in all experiments for both gains and losses and for both health and money.

Delay effect may cause preference reversal as Chapman (2003) suggested with the example of a person who may prefer a large headache relief 8 years from now to a small relief six years from now. However, six years later that person may prefer a small headache relief right away to a large headache relief in the next two years. This indicates that the person discounts the outcome in the first question, which has a longer delay (8 years) less than the outcome in the second question, which has a shorter delay (two years) although both questions involve a two-year delay in order to receive the larger relief. Christensen-Szalanski (1984) also identified a preference reversal in women during childbirth. One month before the labor and during early labor, 18 pregnant women who were interviewed preferred to avoid anesthesia during the labor. Anesthesia had been known as an immediate pain relief but also presenting a risk of long-term side effects. However, during active labor, they shifted their preference to using anesthesia to avoid pain. Thus, in the long delay (one month before labor), the women discounted the long-term side effects of anesthesia less than in the short delay (during active labor). This showed an evidence of a preference reversal as a result of delay effect.

Magnitude effect

Another factor that affects how people discount future outcomes is magnitude of the outcomes. Smaller outcomes were found to be discounted at a higher rate than larger outcomes. Chapman and Elstein (1995) found an evidence of magnitude effect in their time discounting experiments for both health and money domains. The magnitudes of monetary outcomes were $200, $100, $5000, or $25000 gain in the first experiment, and $500, $1000, $2000, $4000 gain in the second experiment. For the health domain, the magnitudes used were six months, one year, two years, or four years in full health in the first experiment, and one year, two years, four years, or 8 years in full health in the second experiment. The results from both experiments showed that the discount rates for both health and money domains were higher when the magnitude of outcome was smaller.

The findings from another study by Chapman (1996b) also support the magnitude effect. All of the results from her three experiments showed that the discount rates for health and money were smaller when the magnitudes of the outcomes were larger. This finding applied to both gains and losses. The magnitudes of outcomes used in her study were $500, $1000, $2000, or $4000 for the money domain, and full health for one year, two years, four years, or eight years for the health domain in the first experiment. In the second and third experiments, the outcomes used were gains or losses for $500, $1500, or $4500, full health or poor health for one year, three years, or nine years. Results showed a magnitude effect in which subjects discounted the outcomes in the future more for smaller outcomes than larger outcomes.

Sign Effect

Discount rates for losses are found to be lower than discount rates for gains. MacKeigan, Larson, Draugalis, Bootman, and Burns (1993) investigated whether people discount future health gains differently from future health losses. A hundred and eight subjects were asked to rate their preference for a hypothetical health gain or loss in 12 health profiles involving the condition of arthritis. The health gains were framed as an arthritis condition followed by excellent health and vice versa for the health losses. Those 12 profiles differed form each other in terms of duration (two days, two months or three years) and delay (one week, two months, or one year). They found that delayed gains were discounted more than delayed losses and the discounting of a delayed loss depended on the duration of the loss. Health losses with long duration showed positive discount rates, but health losses with short duration showed negative discount rate. The fleeting loss was perceived to be worse when delayed. The overall results showed that discount rates were lower for losses than for gains, offering evidence for a "sign" effect. Another support for the sign effect is the finding from a study by Chapman (1996b): she employed gains and losses in both health and money domains and showed that health discount rates in losses were higher than those in gains. However, the results from the money domain were reversed with the money discount rates in gains found to be higher than those in losses. This finding supports the domain dependence in time discounting.

Time Preference and Preventive Health Behavior

Chapman, Brewer, Coups, Brownlee, Leventhal, and Leventhal (2001) explored the relationship between time preferences measured with hypothetical scenarios and actual preventive health behaviors in order to understand whether assessments of time preference with hypothetical scenarios are representative of individuals' real behavior. Three preventive health behaviors that were studied included getting an influenza vaccination, taking medication to control high blood pressure, and taking medication to control high cholesterol. Basically, this study examined whether people who reported a lower discount rate in the hypothetical scenarios would be more likely to engage in the preventive health behavior. The results from all of the three studies showed that there was no relationship between time preferences measured from hypothetical health scenarios and their real-world preventive health behaviors, contrary to the authors' expectations. However, the findings were consistent with those from past studies (Fuchs, 1982; Chapman and Coups, 1999; Chapman, 1998).

Chapman and Coups (1999) examined the relationship between time preference responses in health domain and acceptance of an influenza vaccine. The responses were from 412 corporate employees who were offered free influenza vaccinations at their workplace. Time preferences were assessed for both single outcome (having a flu for two days now or having it start three months from now but stays longer) and sequence of outcomes (having a cough with increased or decreased severity over time for 3 months). For the time preference of the flu, 85% of the responses showed zero time preference. In the sequence of outcomes question, decreasing severity sequences were preferred to increasing severity sequences, indicating a negative time preference for sequence of outcomes. Moreover, the study of the relationship between vaccine acceptance and time preference for the flu vaccination suggested that there was no association between hypothetical health choice-time preference measures and actual preventive health behaviors.

Berndsen and van der Pligt (2001) investigated the role of optimism on time preferences for health gains and losses. Optimism in their study referred to two beliefs about gains and losses: belief that losses will be followed by future opportunities to make up for or avoid these losses (Shelley, 1994) and gains are expected to be followed by additional gains in the future (Chapman and Elstein, 1995). Eighty-one subjects were participated in the study. Each of them was presented with a hypothetical scenario of health gain or health loss in different duration (one year, two years, or four years) and delay (one year, two years, or four years). Health gain was presented as a state with headache followed by a treatment that would lead to perfect health for some duration; whereas health loss was framed as being in a perfect health and followed by headache for some duration. Degree of optimism in the scenario was arranged in two levels: low and high. It was found that low optimism resulted in reduced discount rates compared to discount rates for high optimism for both gains and losses.

2.4.2 Issues in the Studies of Health Time Preference

In the studies of health time preferences, a key challenge is to isolate pure time preferences from other effects. As argued by Gafni (1995), when measuring time preference, effects other than attitude toward timing of the events such as sequence effects are usually confounded with the timing effect. Gafni suggested that when evaluating lifetime health profiles, timing of the event and the sequence of the events are two discrete concepts. For example, when a subject is asked to assess time preference by finding a number of years in a better health state in the future that would make the health profile equivalent to having a specified number of years in that health state now, the response to the assessment does not only measure time preference as intended but instead it incorporates two attitudes: attitude toward the timing of the event (time preference) and attitude toward the sequence of the events (whether he would have good health followed by poor health or poor health followed by good health).

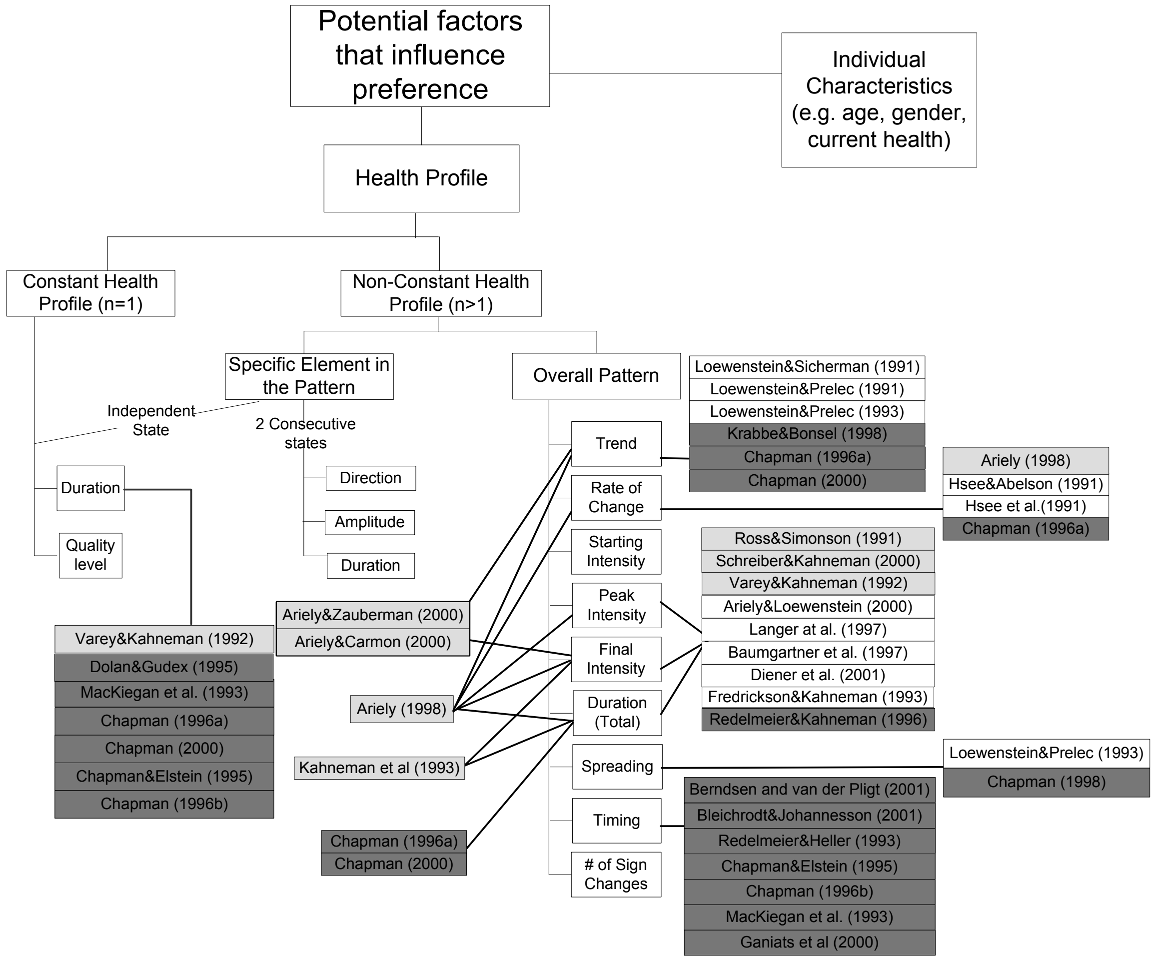
Another issue of time discounting appears in the time-tradeoff (TTO) methodology (Spencer, 2003; Chapman, 2003). Time plays a crucial role in this assessment technique. Thus, timing effect might be confounded with the health outcome preference score as assessed by the TTO technique. The TTO technique elicits the preference for the health outcome by asking individuals to consider two health profiles: poorer health with longer life expectancy and better health with shorter life expectancy. The life expectancy will be adjusted until the individuals are indifferent between the two health profiles. Thus, the two profiles involve two different timings: for example, an individual is indifferent between five years in full health and ten years in poor health. Additional years of life expectancy in the poor health are added to the end of life, which should be considered as a delayed outcome. Later years might be valued less than the earlier ones because of time discounting. The assessment score thus may not be a good indicator of the individual's preference to that particular health state since the time in year 6 to 10 might be discounted.

2.5 Summary

In summary, Figure 2.2 represents the framework of the factors that potentially influence preference for health profiles together with the existing literature that investigated those factors. Factors that drive the preference for the profiles are definitely the health profiles themselves and potentially individual characteristics such as age, gender, etc. Within the scope of this dissertation, effect of individual characteristics will be disregarded. I classify the approach to look at the factors of the health profiles into two categories: when the health profiles are constant (only one health state in the profile) and when the health profiles are non-constant (more than one health state in the profile). In the case of constant health profiles, factors that certainly impact preference are duration and the quality level of the health, identical to the concept of the conventional QALY model. Regarding non-constant health profiles, factors that might impact the preference can be factors from the overall pattern of the profile or factors from specific elements in the profile. Suggested factors from the overall pattern of the profile are trend (increasing or decreasing pattern), rate of change (change of outcome over time), starting intensity, peak intensity, final intensity, total duration of the profile, spreading of the outcomes over time, timing of the outcomes, and number of sign changes (number of times that the profile changes from getting increased to getting decreased and vice versa). When looking at specific elements of the non-constant health profiles, I suggested two approaches: looking at each health state in the profile independently and looking at the two consecutive health states at a time. The first approach is the same as the case of the constant health profile. While in the second approach, the relationship between two consecutive health states is of interest. Factors that might impact this relationship are direction (increasing or decreasing), amplitude (degree of change from one health state to the next one), and duration.

By looking at the list of studies that are corresponding to each factor in Figure 2.2, it is clear that one of the areas that has not yet been explored by any existing study is to look at factors of the relationship between two consecutive health states that might impact the preference of the health profile. Thus, this dissertation will aim at exploring those factors.

Additionally, Table 2.1 provides a summary of the literature discussed in this chapter that addresses all factors that potentially affect preferences for health profiles.



*The dark grey highlights indicate that the domains in the studies were in the health contexts. The light grey highlights indicate the domains in the studies were related to pain or discomfort (e.g. annoying sound, watching unpleasant video clips).

Figure 2.2: Overall picture of the factors that potentially influence preference for health profiles

Table 2.1: Summary of the review of literature

Authors	Domain	Real or Hypothetical Experiment	Findings
Testing the additive independence assumption			
Richardson et al. (1996)	Post mastectomy health profile	Hypothetical	Additive independence assumption was violated.
Kupperman et al. (1997)	Prenatal diagnosis choices	Hypothetical	Additive independence assumption was violated.
Mackeigan et al. (1999)	Treatment profiles for Type 2 diabetes	Hypothetical	Additive independence assumption held. The author explained that the health states in the experiment were too similar.
Spencer (2003)	EQ-5D health scenarios	Hypothetical	Additive independence assumption was violated.
Rate of change (velocity)			
Hsee at al. (1991)	Class rank, stock price, gamble	Hypothetical	Velocity and position of outcome significantly impacted satisfaction.
Hsee and Abelson (1991)	Salary profile, Class grade profile	Hypothetical	Relative importance of position and velocity varied, depending on the condition and nature of the outcome.
Chapman (1996)	Health and money profiles	Hypothetical	Gradually increasing or decreasing sequences were preferred to those that changed steeply.
Ariely (1998)	Pain from heat or mechanical pressure	Real	Rate of change was found as one of the factors that impacted the overall pain evaluation.
Peak, final intensity and duration			
Varey and Kahneman (1992)	Pain, discomfort	Hypothetical	Peak and final intensity significantly impacted the experience evaluation while duration is relatively neglected.
Kahneman et al. (1993)	Discomfort from cold water	Real	Peak and final intensity significantly impacted the experience evaluation while duration is relatively neglected.
Fredrickson and Kahneman (1993)	Watching pleasant and unpleasant video clips	Real	Peak and final intensity significantly impacted the experience evaluation with little effect from duration.
Redelmeier and Kahneman (1996)	Pain in colonoscopy and lithotripsy	Real	Peak and final intensity significantly impacted the experience evaluation with no effect from duration.
Ariely and Zauberman (2000)	Annoying sound	Real	Final intensity significantly impacted the subjects' ratings.
Ariely (1998)	Pain from heat or mechanical pressure	Real	Peak intensity was found to have significant effect but much smaller than the slope and the final intensity.
Ariely and Carmon (2000)	Pain from bone marrow transplant	Real	Final intensity and trend were significant predictors for overall evaluation.

Table 2.1 (continued): Summary of the review of literature

Authors	Domain	Real or Hypothetical Experiment	Findings
Spreading of outcome			
Loewenstein and Prelec (1993)	Two free dinners	Hypothetical	Subject preferred to distribute two free dinners across the time interval.
Chapman (1998)	Money, dinner, health-related events	Hypothetical	Subjects preferred to distribute two events to the 1st and 3rd weekends over a four-week period.
Health Discounting			
Bleichrodt and Johannesson (2001)	Back pain	Hypothetical	Time discounting rates were decreased when the delays were getting longer.
Ganiats et al. (2000)	Five different diseases: chickenpox, Parkinson's disease, tropical disease, migraine headache, and sterilization	Hypothetical	Time discount rates were very high (up to 116%) and they varied among disease conditions.
Redelmeier and Heller (1993)	Health states	Hypothetical	Time discounting rates were decreased when the delays were getting longer.
Chapman and Elstein (1995)	Health and money profiles	Hypothetical	Discount rates for both health and money were decreased as the length of delay increased, and they were higher when the magnitude of outcome was smaller.
Chapman (1996b)	Health and money profiles	Hypothetical	Discount rates for both health and money were decreased as the length of delay increased, and they were higher when the magnitude of outcome was smaller. Moreover, in health, the discount rates in losses were higher than those in gains. However, in money, the discount rates in gains were higher than those in losses.
Mackeigan et al. (1993)	Arthritis condition	Hypothetical	Delayed gains were discounted more than delayed losses.
Chapman et al. (2001)	Preventive health behavior	Hypothetical	There was no relationship between time discount rates measured from hypothetical health scenarios and their real-world preventive health behaviors.
Chapman and Coups (1999)	Preventive health behavior	Hypothetical	There was no relationship between hypothetical health choice-time preference measures and preventive health behaviors.
Berndsen and van der Pligt (2001)	Health gains or losses	Hypothetical	Low optimism resulted in reduced discount rates compared to high optimism for both gains and losses.

Table 2.1 (continued): Summary of the review of literature

Authors	Domain	Real or Hypothetical Experiment	Findings
Trend			
Loewenstein and Prelec (1991)	Dinner at a French or Greek restaurant	Hypothetical	Subject preferred improving sequences.
Loewenstein and Sicherman (1991)	Wage profiles	Hypothetical	Subject preferred improving sequences of wages.
Loewenstein and Prelec (1993)	Free dinner	Hypothetical	Subject preferred to distribute two free dinners across the time interval.
Chapman (1996a)	Health and money profiles	Hypothetical	Subjects preferred improving sequences for both health and money for short sequences. For long sequences, subjects preferred decreasing sequence for health but increasing sequence of money.
Chapman (2000)	Headache pain, athletic ability, facial acne, wrinkles	Hypothetical	Subjects strongly preferred improving sequences.
Ariely and Zauberman (2000)	Annoying sound	Real	Subjects rated pattern that had final increasing trends as more annoying than those that had final decreasing trends.
Ariely and Carmon (2000)	Pain from bone marrow transplant	Real	Final intensity and trend were significant predictors for overall evaluation.
Krabbe and Bonsel (1998)	EQ-5D health scenarios	Hypothetical	The results evidenced for sequence effect.

CHAPTER 3

RESEARCH DESIGN

3.1 Problem Statement

In preference assessment for multistate health profiles, the existing literature has shown that preferences for entire health profiles are potentially impacted by many factors such as trend (Krabbe and Bonsel, 1998; Chapman, 1996a, 2000), rate of change (Chapman, 1996a; Ariely, 1998), peak and final intensity (Ariely, 1998; Ariely and Carmon, 2000; Varey and Kahneman, 1992; Redelmeier and Kahneman, 1996; Ross and Simonson, 1991), spreading of outcomes (Chapman, 1998), and timing of events (Bleichrodt and Joahnnesson, 2001; Redelmeier and Heller, 1993; Chapman and Elstein, 1995; Chapman, 1996b; MacKeigan et al., 1993). Thus, as the number of health states in the profile becomes large, the preference assessment of the entire health profile becomes highly complex. If those factors are not taken explicitly into account, the preference scores derived might not accurately represent the subjects' preferences.

In order to simplify the task, decomposition of the assessments may be employed. One decomposition method that has been used is to decompose the multistate health profile into a series of independent health states. Each health state is then evaluated independently and all are combined to produce the overall evaluation of the profile. However, as the past literature showed, decomposing by evaluating each constituent state independently cannot accurately predict the overall evaluation of the profile since the preference for health states are not independent to each other (Spencer, 2003; Richardson et al., 1996; Kupperman et al., 1997). Moreover, different patterns of the health profiles have been found to impact individuals' preferences.

In this dissertation, I propose a novel approach to decompose the preference assessment task for the entire multistate health profile. The concept is based on the findings that preference from each health state is not independent of other health states in the profile and that health profile patterns have been found to be a major factor. By looking at key characteristics of the health profile patterns, say, looking at two consecutive health states at a time, the nature and extent of the interdependence between consecutive health states is expected to be revealed and used in assessing the overall profile. The rationale in looking at the interdependence between two consecutive health states is based upon the suggestion that people use reference points in evaluating the attractiveness of choices. For example, Loewenstien and Sicherman (1991) stated that people prefer an improving sequence as opposed to a declining one since they use their current consumption as a reference point. The future outcomes in the improving sequence are perceived as gains when compared with the previous (or current) state. Moreover, in medical decision making, people need to make decision for future (uncertain) outcomes, for example, making a decision regarding alternative drugs that will improve health states in the near future. In this example, individuals might guide their drug choice by comparing the future health outcome they expect as a result of taking the drug with the current health state they are experiencing. Thus, it is plausible that the current health state constitutes a reference point for evaluating future health outcomes. This example also illustrates that there is a potential tendency for the interdependence between two consecutive health states. Another rationale is that human beings obviously experience health states in a chronological order. In their lives, health states change over time, with different states following one another. As a result, it is feasible that there is an interdependent relationship between consecutive health states. Thus, I hypothesize that current health state will have an effect on the preference of the following health state, which will have an effect on the preference of the next following state, and so on. For example, a subject may evaluate the next health state, say health state X, as pleasant if his/her current health state is worse than health state X. On the other hand, the same subject may evaluate the same health state X as an unpleasant state if his/her current health state is better than health state X. Thus, decomposing the assessment of the entire health profile into several "conditional preference score" assessments by looking at two consecutive states at a time is expected to better predict the preference score for the entire health profile than looking at each health state independently since the interdependent relationships between health states would be taken into account.

Definitions:

1. *A "conditional preference score" of the future health state is the term that I use for a preference score that is assessed by considering the relationship of the current health state to the following health state. For example, a conditional preference score for health state B in Figure 3.1 is obtained by assessing a preference score of health state B given that one is currently in health state A. A conditional preference score for health state C is obtained by assessing a preference score of health state C given that one is currently in health state B.*

Note that it is an evaluation of the later health state and not the evaluation of the transition from one state to the next state.

2. *A "future" health state in the context of this study refers to a future health state that immediately follows the current health state. If the current health state is at time "n", the "future" health state in this context means the health state in time "n+1" only, not including the states beyond state "n+1".*

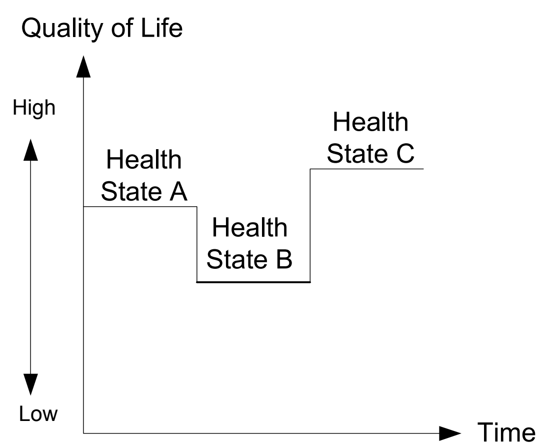


Figure 3.1: A health profile composed of three health states, health state A, B, and C, sequentially

3.2 Significance of the Study

A major contribution I am looking forward to obtaining from this study is an improved health outcome measurement system for multistate health profiles. The existing method in which each constituent health state is assessed independently cannot predict the preference for entire health profile as the literature has shown (Spencer, 2003; Richardson et al, 1996; Kupperman et al., 1997). In cost effectiveness analysis, an economic analysis technique used in health policy, health outcome measurement plays a critical role that impacts the medical decision making process. Whether scarce resources in healthcare are distributed appropriately depends on how accurate health outcome measurements are. I hope that this study will result in a better health outcome measurement system, which consequently will result in more accurate cost effectiveness analyses, which in turn will contribute to better healthcare resource allocations.

Moreover, at a more specific level, exploring the interdependent relationship between two consecutive health states will assist in understanding how different factors influence overall health profile preference assessment. Those factors include peak intensity, final intensity, trend, rate of change, etc. Furthermore, this study will shed some light on how combinations of factors jointly influence preferences for health profiles. For example, the amplitude between two consecutive states in the health profile also contributes to the rate of change or "velocity" (as defined by Hsee and Abelson, 1991). High amplitude between two consecutive states would contribute to an increase in the velocity. If it is a gain, high amplitude will contribute to a more positive velocity. If it is a loss, high amplitude will contribute to a more negative velocity. Lastly, in the research perspective, I hope that this study will inspire further studies in exploring the relationship between consecutive health states. For example, further studies can explore the interdependence among three consecutive health states.

3.3 Research Questions

The following two main research questions have been formulated:

1. Is a preference score for future health state dependent upon level of current health state?
2. Can "conditional preference scores" for discrete health states better predict preference scores for an entire health profile than unconditional health states assessments?

The methods section, which will be described later, is divided into two phases. The design in Phase 1 aims at answering the first research question while Phase 2 aims at answering the second research question.

3.4 Methods - Phase 1

In order to ensure that the conditional preference assessment technique that I propose will yield different result from the existing decomposition technique in which each health state is evaluated independently, the initial step is to explore if the relationship between two consecutive health states really exists, or explicitly, if a preference score for a future health state is dependent upon the level of the current health state and if so, what factors related to the current health state affect the "conditional preference score" for the future health state. As a result of the literature review, three potential factors are investigated in this study: direction of change from current health state (whether it is considered as gain or loss), amplitude of change from current to future health state, duration that an individual has been in the current health state, and two- and three-way interactions among the three factors.

3.4.1 Study Factors

Factor#1: Direction of change from current health state

People might evaluate the same health state differently depending on whether that health state is perceived as gain or loss. Direction of change from current health state to future health state contributes to the perceived appearance of the future health state. If the future health state is increased from the current one, it will be perceived as a gain, and vice versa. People might overvalue the preference score when the future health state is perceived as a gain and undervalue the preference score when it is perceived as a loss. For example, see Figure 3.2, health state X is an individual's future health state. Health state A and B are current health states. Health states X in A) and B) are at the same level, however, in A), the current health state (health state A) is higher than health state X, while in B), the current health state (health state B) is lower than health state X. Thus, health state X in A) is considered as a loss while health state X in B) is considered as a gain. The question is if an individual's future health state is health state X and he/she is asked to assess his/her preference score for health state X, will he/she give the same score for health state X if health state X is considered as a loss (A) or as a gain (B)?

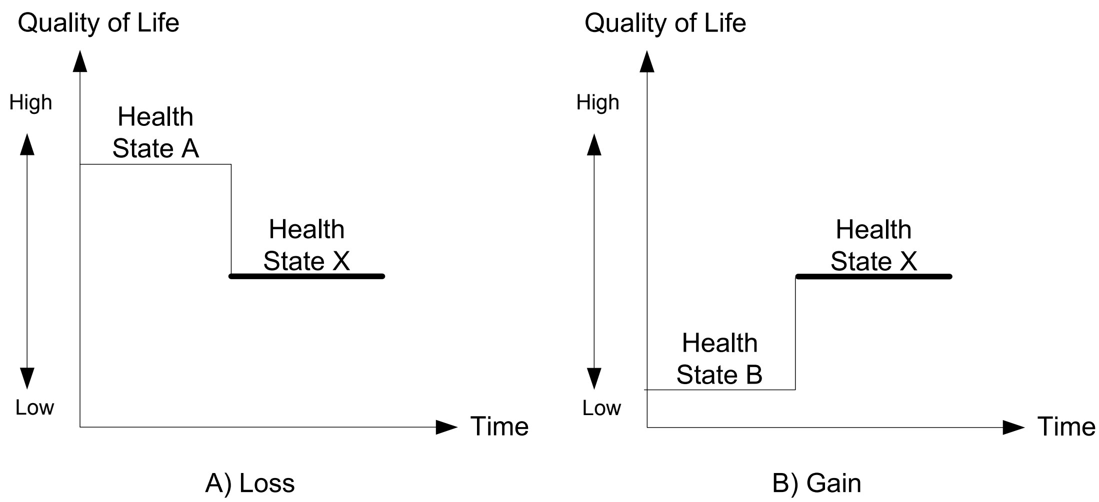


Figure 3.2: Illustration of potential effect from direction of change from current health state

Note that this concept is slightly different from the direction effect reviewed in the existing literature since the direction effect related to whether people prefer improving or declining profiles. However, in this one, I refer to the effect of direction from one current health state on an evaluation of a future health state (not the entire health profile).

<u>Factor #2: Amplitude of change from current health state to future health state</u>

Amplitude of change from the current health state to the future health state is suspected to affect the conditional preference score of the future health state since when people are asked to assess their preference scores for their future health states, they might make an evaluation by comparing their future health states with the health state they currently are in and make an evaluation based on the difference. Amplitude indicates the absolute change between two health states (e.g. current health state and future health state, in this case). Different amplitudes may yield different preference scores for the same health state. For example, see Figure 3.3, health state X is an individual's future health state. Health state A and B are current health states. Health states X in A) and B) are at the same level, however, the amplitude of change from health state A to health state X is lower than the amplitude of change from health state B to health state X. The question is if an individual is asked to value a future health state, which is health state X in Figure 3.3, will he/she give the same score for health state X if the health state that he/she is currently in is different (health state A or health state B)?

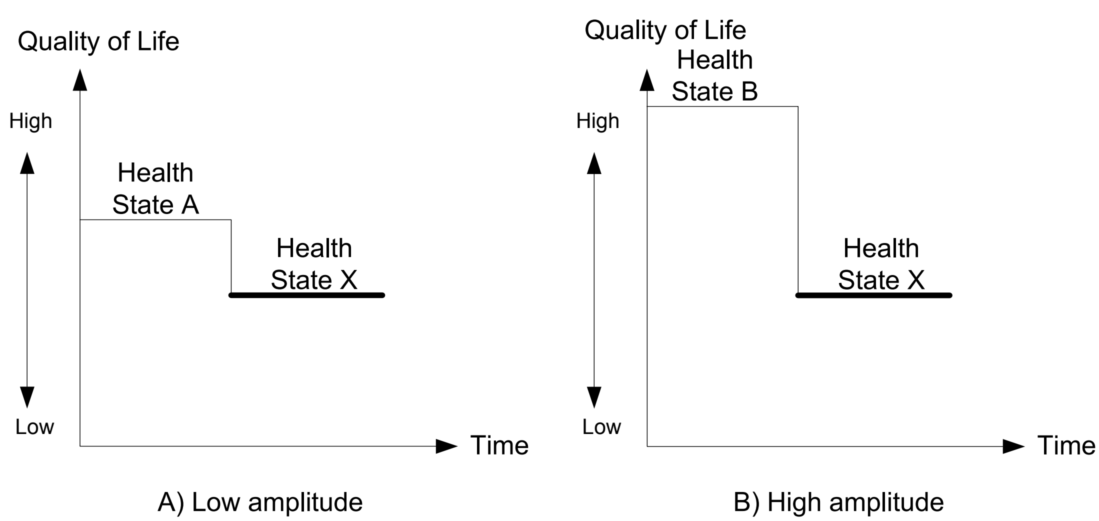


Figure 3.3: Illustration of potential effect from amplitude of change from current health state to future health state

Factor#3: Current health state duration

Duration that an individual has been in his/her current health state might impact the degree of interdependence between the two consecutive health states, which consequently impact the conditional evaluation of the future health state. For example, see Figure 3.4, health state X is an individual's future health state. Health state A and B are current health states. Health states X in A) and B) are at the same level. However, the current health state, health state A in A) has shorter duration than the current health state, health state B in B). The question is if an individual is asked to value his/her future health state, health state X in Figure 3.4, will he/she give the same score for health state X if he/she has been staying in his/her current health state in 1 year (health state A) or in 10 years (health state B)?

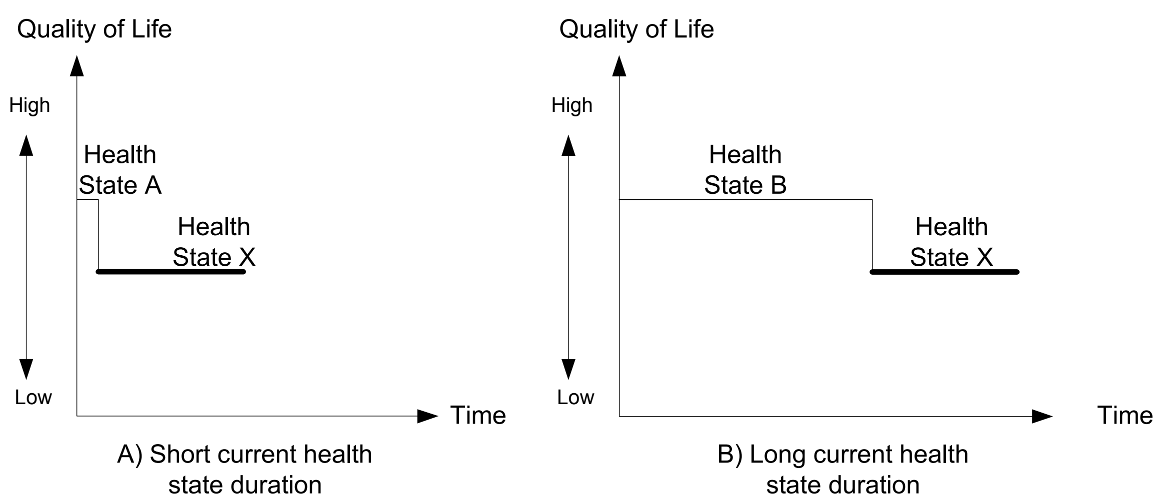


Figure 3.4: Illustration of potential effect from current health state duration

Duration in current health state is considered as one of the factors since I suspect that duration may relate to how people value their current health state. If people value their current health state differently as a result of duration, then that would affect the relative difference between current health state and future health state which might consequently affect the conditional preference evaluation of the future health state if the interdependence between two consecutive health states actually exists. Moreover, duration in the current health state may relate to how well people remember and use the experience in the current health state as their reference point when evaluating their future health state.

3.4.2 Responses

Responses for the study in Phase 1 are conditional preference scores for future health states in different hypothetical health scenarios. The hypothetical health scenarios will be described later in the experimental design section.

3.4.3 Hypotheses

Since the three potential factors (direction, amplitude, and duration) above may interact with each other, I thus start the hypotheses by first looking at the interactions, then the main effects.

Hypothesis 1: There is no significant three-way interaction effect between direction, amplitude and duration on the conditional preference score of the future health state.

Hypothesis 2: There is a significant effect of interaction between direction and amplitude. See Figure 3.5, health state X is an individual's future health state. Health state A, B, C and D in each graph is his/her current health state. Health states X in A) and B) are at the same level. Health states X in A and B are perceived as losses; however, A has lower amplitude of change from current health state than in B. Health states X in C and D are perceived as gains; however, C has lower amplitude of change from current health state than D. If future health state is perceived as a gain (increased from current health state), people will prefer the future health state more (overvalue conditional preference score) if the amplitude of change from the current health state is larger (Figure 3.5: C and D). On the other hand, if the future health state is perceived as a loss (decreased from current health state), people will prefer the future health state less (undervalue conditional preference score) when the amplitude of change from the current health state is larger (Figure 3.5: A and B).

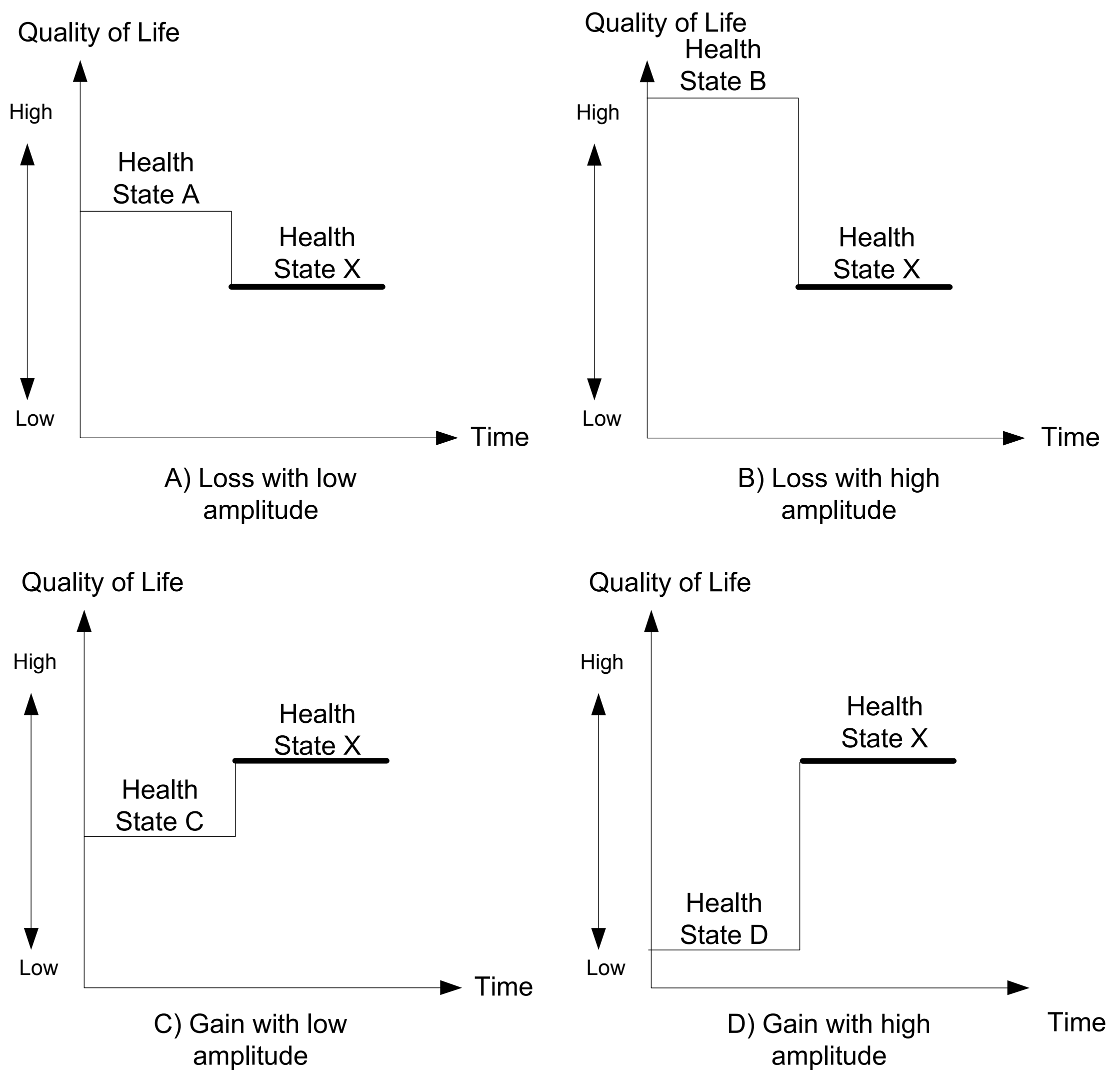


Figure 3.5: Illustration of potential effect of 2-way interaction between direction and amplitude

Hypothesis 3: There is no significant interaction effect between direction and duration on the conditional preference score of the future health state.

Hypothesis 4: There is no significant interaction effect between amplitude and duration on the conditional preference score of the future health state.

Hypothesis 5: A main effect of duration will result in the interdependence between the future health state and the current health state. The effect will be less significant on preference when the current health state has shorter duration. One may argue that if people stay in the poor health state longer, they may be able to accommodate to the state and thus may not value their current health state as poorly as the people who stay in that health state in a shorter duration would. If this argument is true, then the perceived amplitude of change may vary as the duration in current health state varies thus making it hard to control the amplitude factor. However, a study by Myers, McPherson, Taylor, Weatherall and McNaughton (2003) found no evidence that the duration of health state was related to health-state valuation on the EQ-5D self-evaluation instruments. As will be explained later, the EQ-5D system will be used to construct scenarios thus minimizing the risk of the inability to control amplitude change.

Note: Main effects of direction and amplitude can be ignored since in Hypothesis 2, I hypothesize that there is a significant interaction effect between direction and amplitude on conditional preference score of current health state.

3.4.4 Experimental Design

In Phase I, the three factors of interest are direction, amplitude, and duration of current health state. Those factors are varied in order to explore the effect of each factor on the conditional preference score of the future health state as well as interaction effects. Two levels of each factor are used in the experiment, which are as follows:

1. Direction of change from the current health state to the future health state: increase or decrease

- Increase: when the future health state is perceived as gain
- Decrease: the future health state is perceived as loss

2. Amplitude of change: large or small

- Large: The size of amplitude of change is fairly a relative concept. From a health weight scale of 0 (death) to 1 (perfect health). The smallest "large" amplitude that I can use has to be smaller than 0.50 since when using "large" amplitude for increase and decrease directions, the current health states used in the scenarios will be at levels 1.00 and 0.00. Using level 0.00 as a current health state is unreasonable since level 0.00 is equal to death. Thus, I consider a change of 0.30 on the health weight scale to be the "large" amplitude, so that replications of the experiment with different levels of future health states are possible.
- Small: I consider a change of 0.15 on the health weight scale to be the "small" amplitude.

3. Duration of prior health state: long or short

- Long: 10 years
- Short: 1 year

A 2^3 full factorial design (three factors with two levels each) for a total of 8 scenarios are used for each future health state assessed. The response is the conditional preference score for the future health state. The design is shown in Table 3.1. Each scenario consists of two health states: current and future health states. The level of future health state is fixed for all 8 scenarios. However, the current health state is varied between each scenario (four different levels with two different durations: 4 x 2 = 8). Figure 3.6 shows the 8 scenarios corresponding to the design in Table 3.1.

Table 3.1: A 2^3 factorial design layout of three factors: direction, amplitude, and duration. Order of run will be randomized for each subject.

Scenarios	**Direction**	**Amplitude**	**Duration**
1	Decrease	0.30	1 year
2	Decrease	0.30	10 years
3	Decrease	0.15	1 year
4	Decrease	0.15	10 years
5	Increase	0.30	1 year
6	Increase	0.30	10 years
7	Increase	0.15	1 year
8	Increase	0.15	10 years

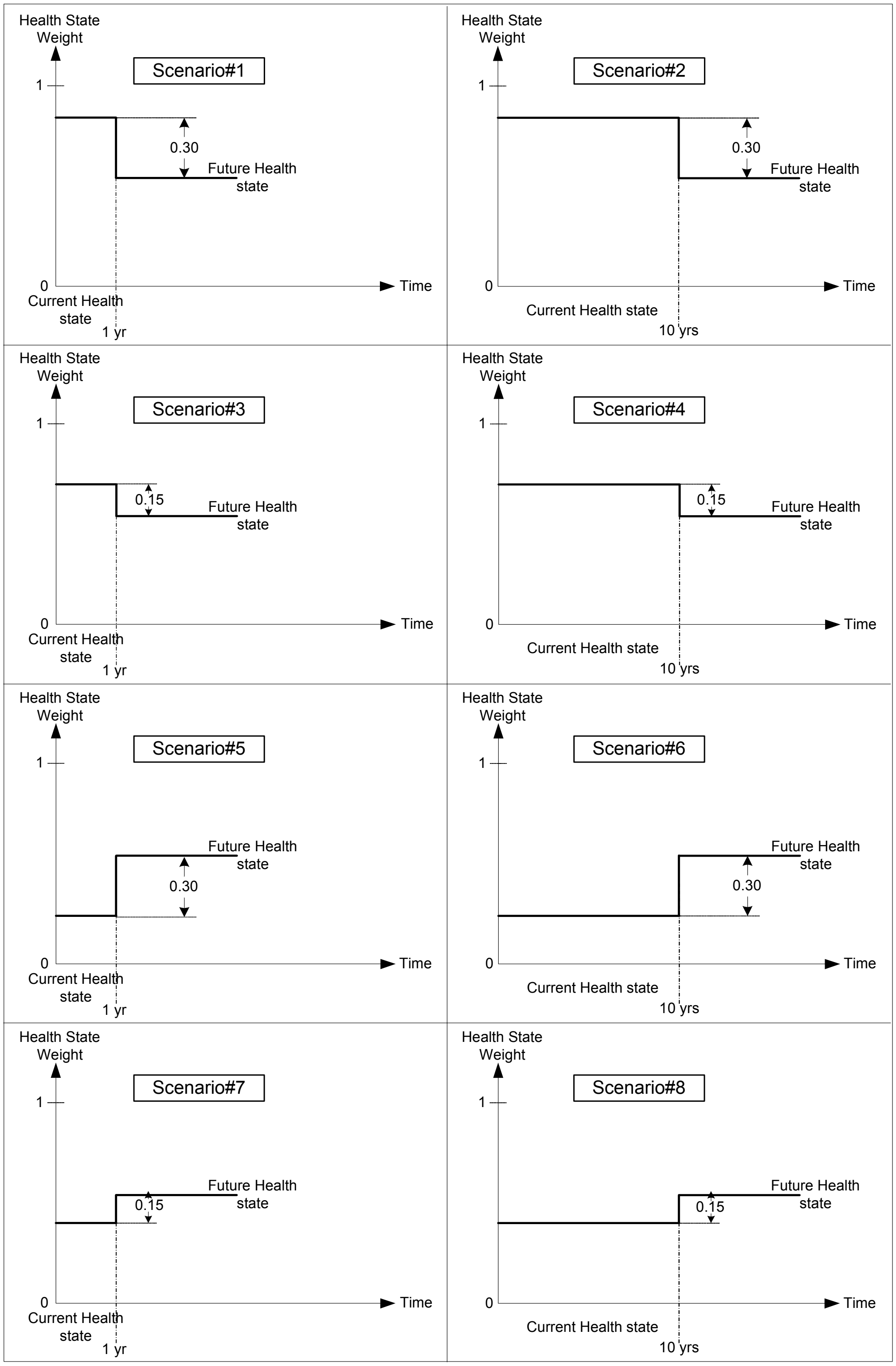
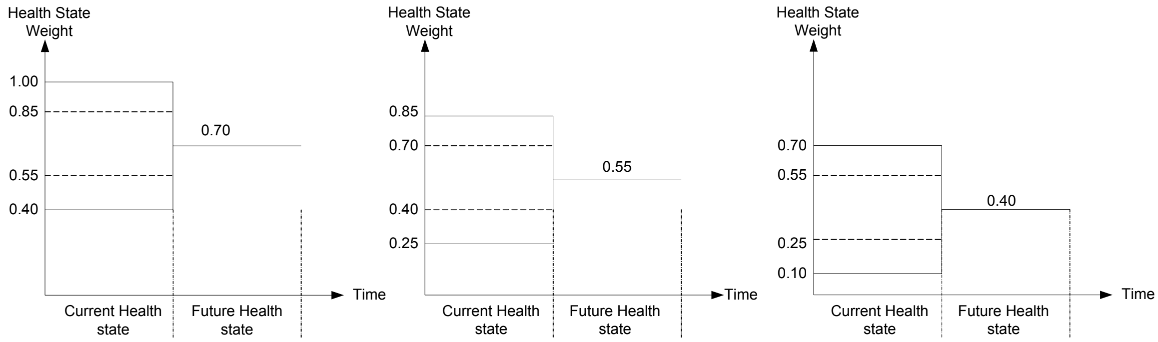


Figure 3.6: Graphical representation of the 8 scenarios from the 2^3 factorial design in Table 3.1

The design in Table 3.1 and Figure 3.6 shows that for each level of future health state examined, 8 scenarios are required. Replications of the experiment with different level of future health states are recommended in order to get the better and more reliable results. Thus, more than one set of the design in Table 3.1 is recommended to test at other levels of future health state. With the limitation of health state weight range of 0 and 1 and the required amplitude range (up to 0.30 for large amplitude) in the experiment, the highest level of future health state that can be examined is 0.70 (1-0.30 = 0.70, the 0.30 tolerance is required so that the current health state does not go beyond 1.00 in the scenario of large amplitude with the decrease direction). In addition, with the same rationale, the lowest level of current health state that can be examined have to be greater than 0.30 (cannot be 0.30 since the current health state cannot be 0 (death)). With these restrictions, a total of three levels of future health state are proposed with levels set at 0.70, 0.55, and 0.40. Figure 3.7 shows the levels of current and future health states that are required in the experiments (different durations are not shown in the Figure). Thus, the design in Table 3.1 is repeated with three different levels of future health states. A total of 24 scenarios will be used to assess three levels of future health states. Hereafter, I will refer to these three sets of the design as Design A, Design B, and Design C for the following levels of the future health state: 0.70, 0.55, and 0.40, respectively. Figure 3.8, Figure 3.9, and Figure 3.10 show the 8 scenarios corresponding to Design A, B, and C in Figure 3.7, correspondingly.



Design A: Future health state is at 0.70 level Design B: Future health state is at 0.55 level Design C: Future health state is at 0.40 level

Figure 3.7: Graphical illustrations of the Design A, B, and C

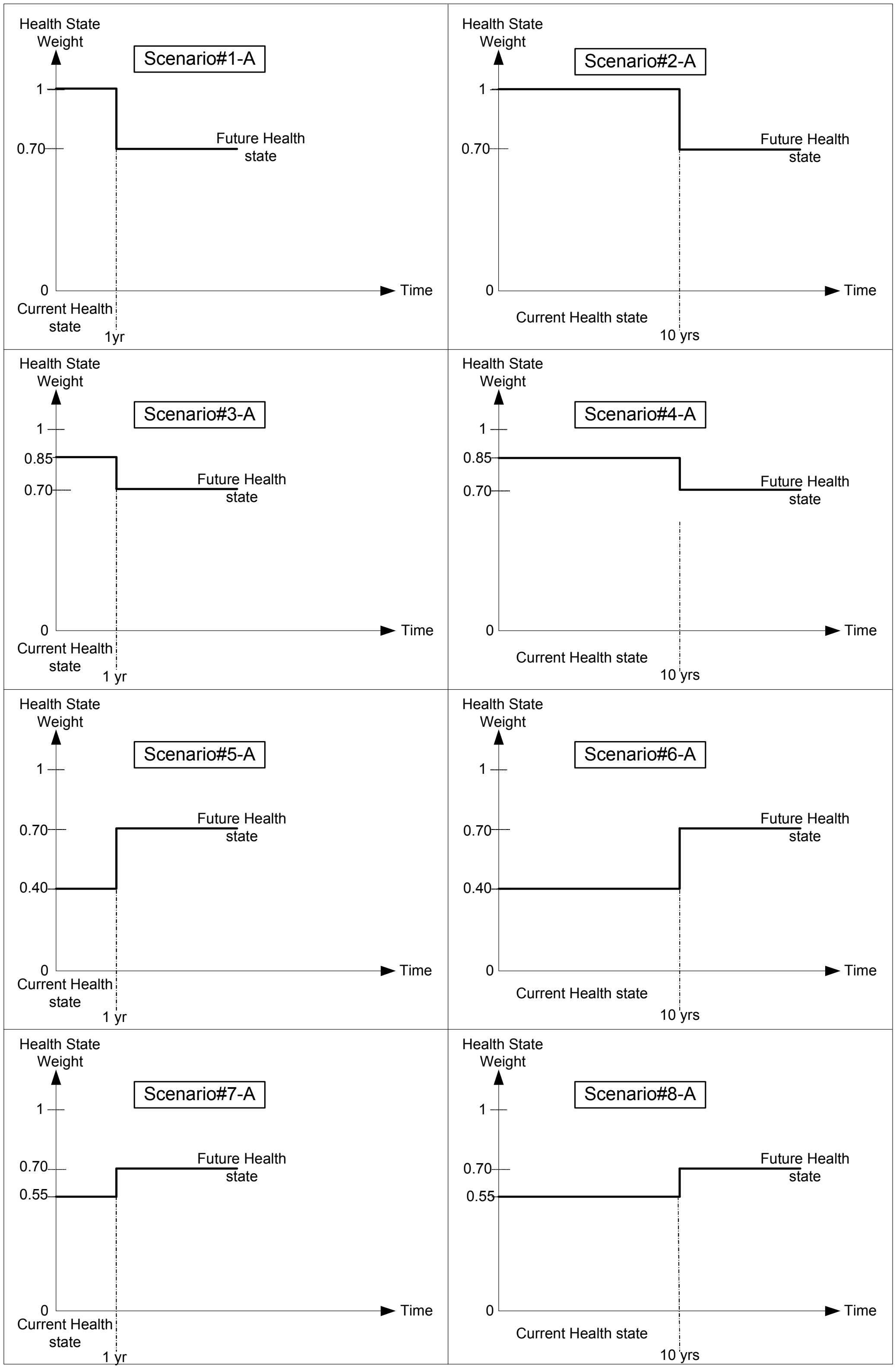


Figure 3.8: Graphical representation of the 8 scenarios for Design A in Figure 3.7

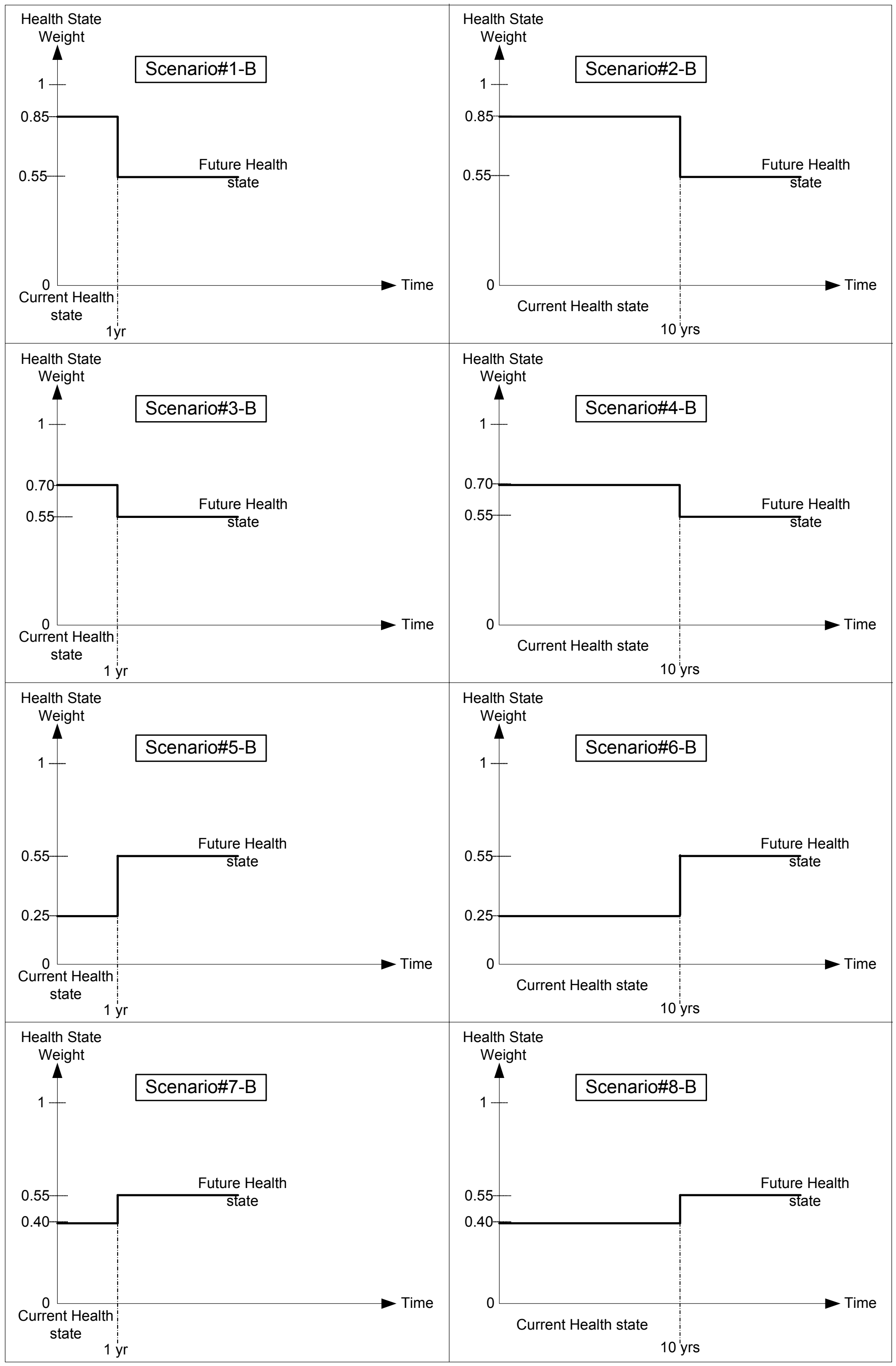


Figure 3.9: Graphical representation of the 8 scenarios for Design B in Figure 3.7

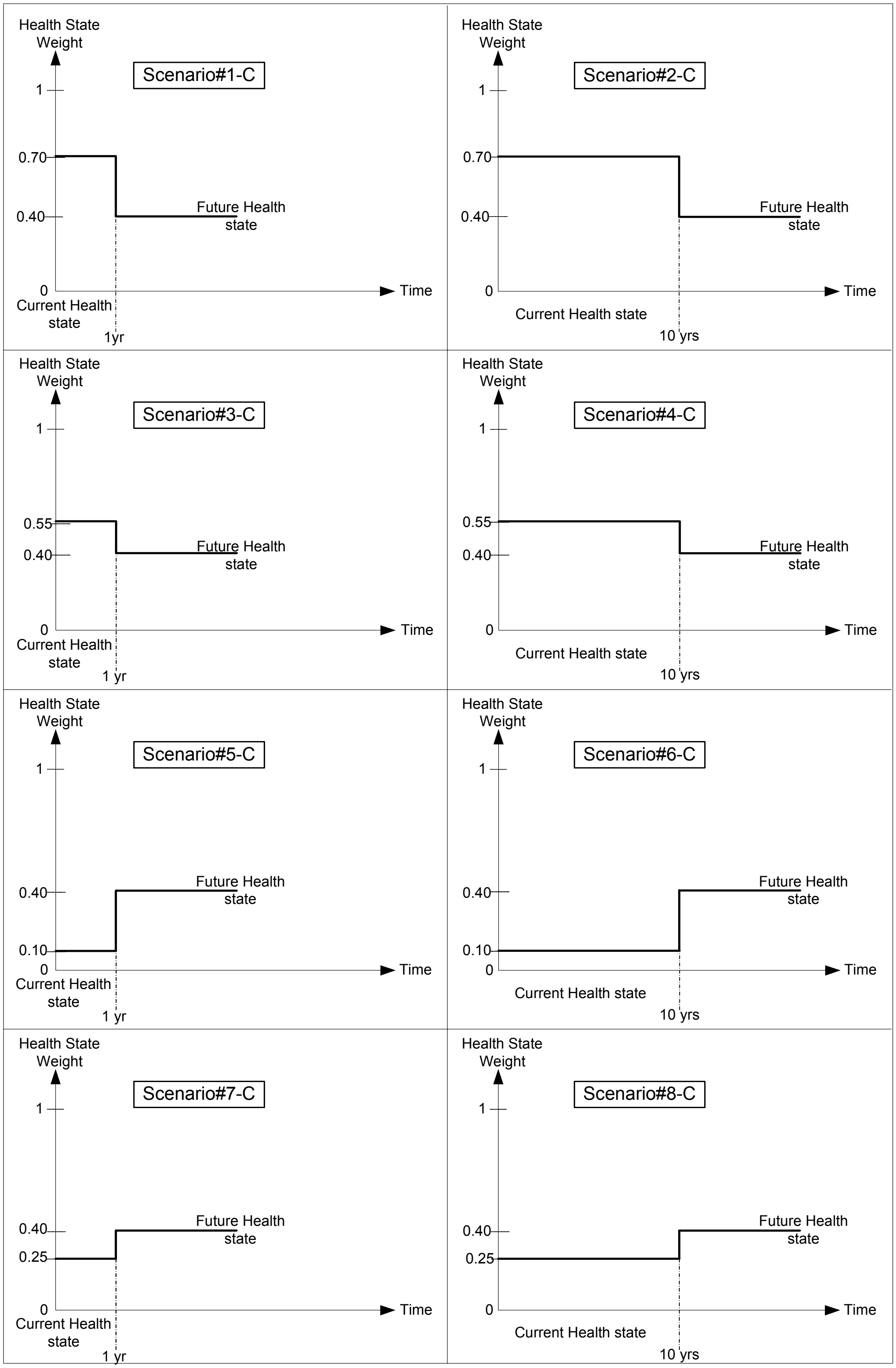


Figure 3.10: Graphical representation of the 8 scenarios for Design C in Figure 3.7

3.5 Methods - Phase 2

The design of experiment in Phase 2 aims at answering the second research question: Can "conditional preference scores" for discrete health states better predict preference scores for an entire health profile than unconditional health states assessments? If the interdependent relationship between two consecutive health states exists, I believe that the proposed decomposition method can better predict the preference score of the entire health profile than the existing decomposition method in which each constituent health state is assessed independently, since the effects of transition between health states on preferences are taken into account.

3.5.1 Study Factor

The only factor of interest in the second research question is the health profile evaluation techniques. Evaluation techniques are employed as a within-subject factor. Three types of evaluation techniques are as follows:

1. Holistic preference score assessment

The holistic assessment can be obtained by presenting the subjects the entire multistate health profile and asking them to evaluate the entire profile. For example, health profile A in Figure 3.11, which consists of four health states: health state 1 in duration a, health state 2 in duration b, health state 3 in duration c, and health state 4 in duration d, will be presented to each subject. Each subject will assign a score that represents how they feel about the health profile.

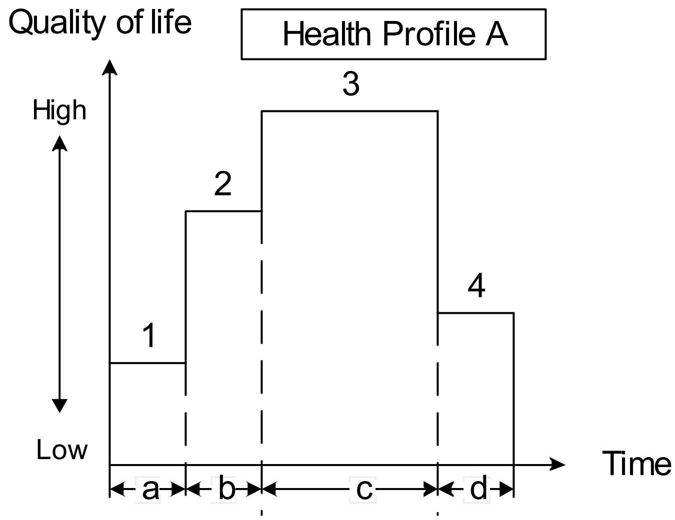


Figure 3.11: Health profile A

2. Unconditional preference score assessment

The unconditional preference score assessment can be obtained by presenting the subjects each health state in the multistate health profile and asking them to evaluate each health state independently. For example, Figure 3.12 conceptually illustrates how the unconditional preference scores for health profile A in Figure 3.11 can be assessed. Note that in unconditional preference assessment for each health state, time does not matter.

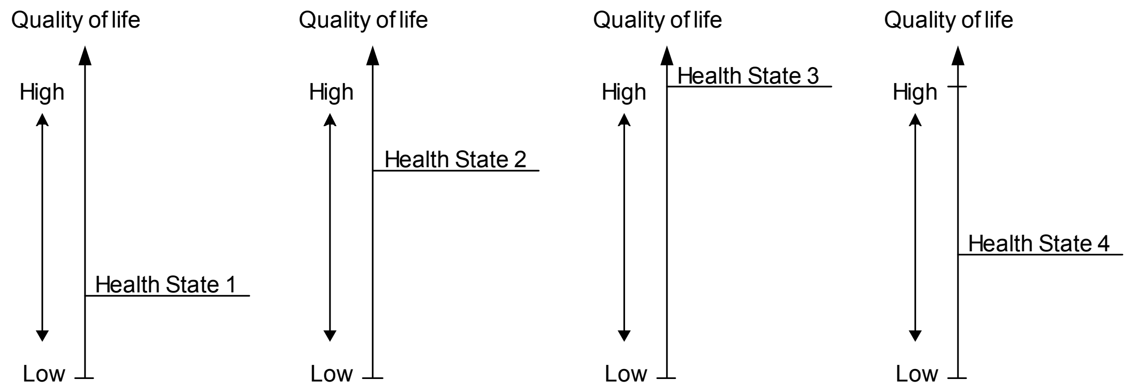


Figure 3.12: Conceptual illustration of unconditional preference score assessment for health profile A in Figure 3.11

3. Conditional preference score assessment

The conditional preference score assessment is the method that is proposed in this dissertation. The conditional preference assessment will be done at each two consecutive health states in the multistate health profile by assessing the conditional preference score for the later health state. For example, health state n+1 can be assessed by presenting state n and state n+1, health state n+2 can be assessed by presenting state n+1 and state n+2, and so on. However, the first state in the health profile will be assessed independently. Thus, the first health state score from the conditional preference assessment and the unconditional preference assessment are identical. Figure 3.13 conceptually illustrates how the conditional preference scores for health profile A in Figure 3.11 can be assessed. Figure 3.13 A, B, C and D represent how to obtain preference scores for health state 1, 2, 3, and 4, respectively.

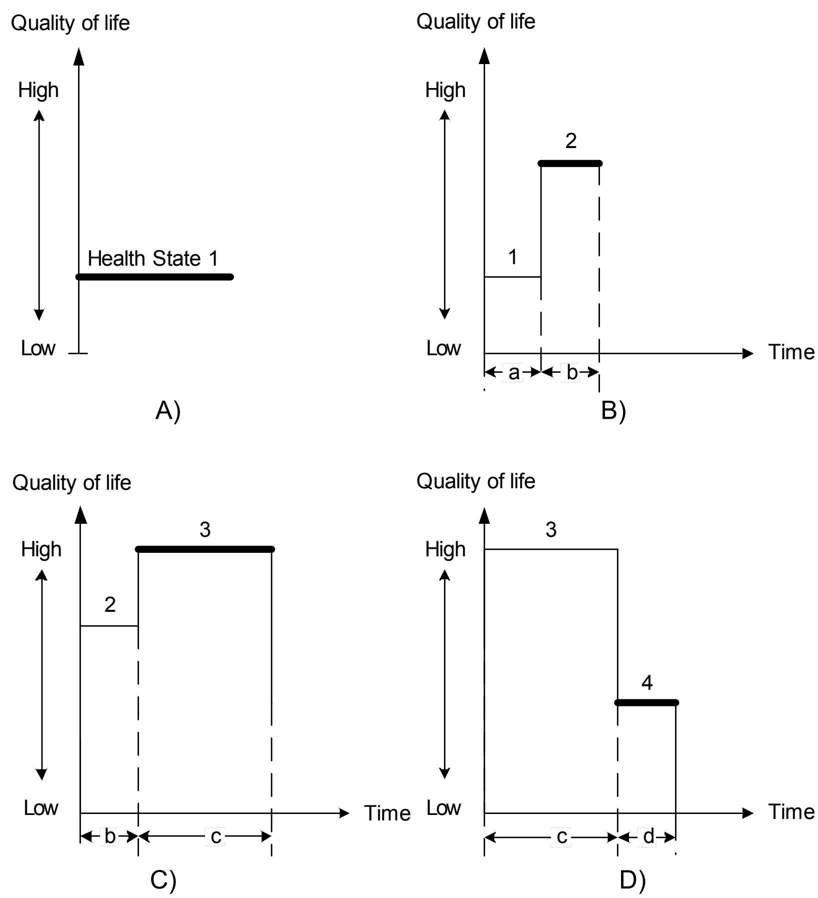


Figure 3.13: Conceptual illustration of conditional preference score assessment for health profile A in Figure 3.11

3.5.2 Responses

Thus, responses in Phase 2 study provide scores from the three assessment techniques described above (e.g. holistic, conditional, and unconditional). After obtaining the scores from the assessments, the scores are duration-weighted in order to generate a single score for the health profile from each technique and these scores then are compared across the assessment techniques. An example of the process in calculating the duration-weighted scores for the health profile A in Figure 3.11 is described as follows:

1. Duration-Weighted Holistic Preference Score = (a+b+c+d) × Score (Profile A)

2. Duration-Weighted Unconditional Preference Score

   = [a × Score(1)] + [b × Score(2)] + [c × Score(3)] + [d × Score(4)]

3. Duration-Weighted Conditional Preference Score

   = [a × Score(1)] + [b × Score(2|1)] + [c × Score(3|2)] + [d × Score(4|3)]

<u>Notations:</u>

Score(Profile A) = Holistic preference score for health profile A

Score(1) = Unconditional preference score for health state 1

Score(2) = Unconditional preference score for health state 2

Score(3) = Unconditional preference score for health state 3

Score(4) = Unconditional preference score for health state 4

Score(2|1) = Conditional preference score for health state 2 (future state) given that the current health state is health state 1

Score(3|2) = Conditional preference score for health state 3 (future state) given that the current health state is health state 2

Score(4|3) = Conditional preference score for health state 4 (future state) given that the current health state is health state 3

3.5.3 Hypothesis

<u>Hypothesis 6:</u> The proposed decomposition method in which several conditional preference score assessments are executed and integrated to produce a preference score for the entire health profile will predict the preference of the entire health profile better than the decomposition technique that uses unconditional preference score assessments (the conventional QALY model).

3.5.4 Experimental Design

The experiment in Phase 2 aims at comparing the holistic preference assessments with the conditional preference score assessments and the unconditional preference score assessments for the same health profiles. In order to minimize the additional evaluation tasks for subjects, health profiles that are used in the Phase 2 experiment are composed of health states that are used in Phase 1. Thus, the health profiles used are composed of health states at levels 0.10, 0.25, 0.40, 0.55, 0.70, 0.85, and 1.00. Moreover, different levels of the three factors interested in Phase 1 (direction, amplitude, and duration) are varied among the health profiles. Notice that when several health states are integrated into a health profile, amplitude of change between each consecutive health states plays a role as another new factor, “rate of change” or “slope”. Larger amplitude between consecutive health states contributes to higher rate of change (or steep slope) of the health profile, and vice versa. Thus, patterns of the health profiles in Phase 2 are constructed from a factorial design by varying three factors: direction (2 levels: increase and decrease), duration of the health profile (2 levels: 4 years and lifetime), and slope (2 levels: gradual and steep). A 2^3 factorial design yields a total of 8 different health profiles. Moreover, two additional health profiles, which have no systematic pattern, one has 4-year duration and another one has lifetime duration, are included. Followings are 10 health profiles that are constructed for Phase 2 experiment. Table 3.2 summarizes key characteristics of each health profile. Additionally, Figure 3.14 to Figure 3.23 demonstrate graphical representations of the 10 health profiles.

Table 3.2: Summary of key components of each health profile used in Phase 2 experiment

Health Profiles#	Duration	Slope	Direction
1	4 Years	Gradual	Decrease
2	4 Years	Gradual	Increase
3	4 Years	Steep	Decrease
4	4 Years	Steep	Increase
5	4 Years	No systematic pattern	
6	Lifetime (40 years)	Gradual	Decrease
7	Lifetime (40 years)	Gradual	Increase
8	Lifetime (40 years)	Steep	Decrease
9	Lifetime (40 years)	Steep	Increase
10	Lifetime (40 years)	No systematic pattern	

Health profile 1 is composed of four health states with one-year duration for each health state. The changes between each two consecutive health states have low amplitude (gradual slope) with decreasing direction.

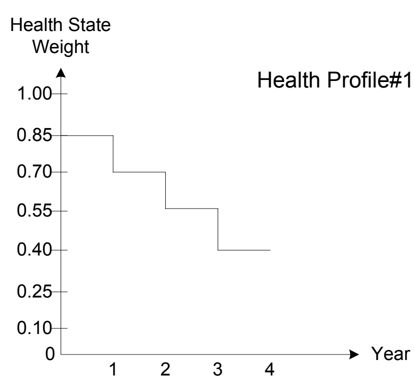


Figure 3.14: Health profile 1 for Phase 2 experiment

Health profile 2 is composed of four health states with one-year duration for each health state. The changes between each two consecutive health states have low amplitude (gradual slope) with increasing direction.

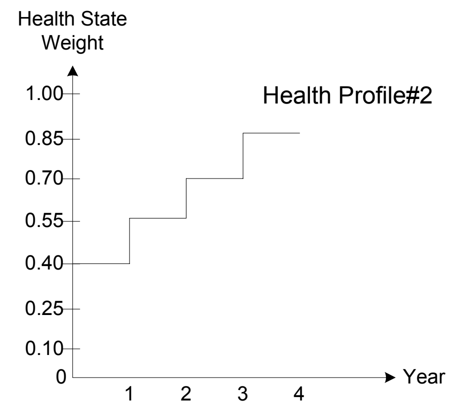


Figure 3.15: Health profile 2 for Phase 2 experiment

Health profile 3 is composed of four health states with one-year duration for each health state. The changes between each two consecutive health states have high amplitude (steep slope) with decreasing direction.

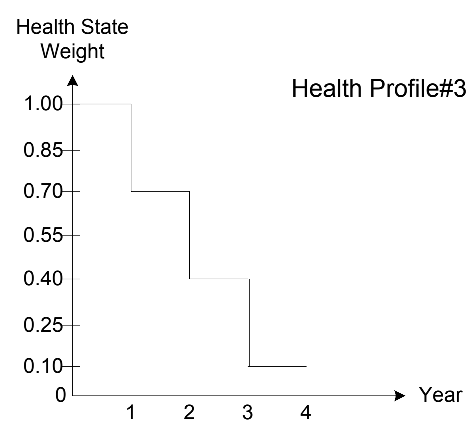


Figure 3.16: Health profile 3 for Phase 2 experiment

Health profile 4 is composed of four health states with one-year duration for each health state. The changes between each two consecutive health states have high amplitude (steep slope) with increasing direction.

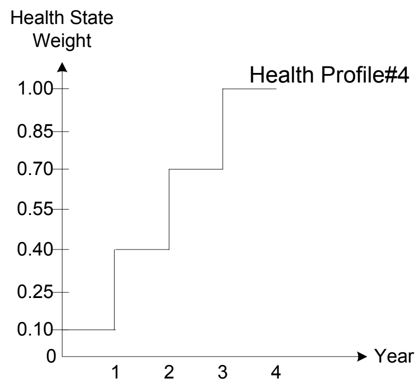


Figure 3.17: Health profile 4 for Phase 2 experiment

Health profile 5 is composed of four health states with one-year duration for each health state. The profile has no systematic pattern. The changes between each two consecutive health states compose of a gain with high amplitude and two losses with high and low amplitudes, respectively.

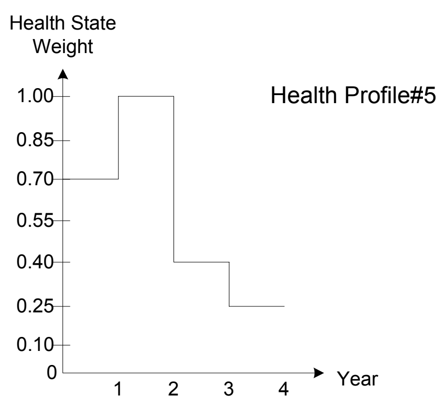


Figure 3.18: Health profile 5 for Phase 2 experiment

Health profile 6 is composed of four health states with 10-year duration for each health state. The changes between each two consecutive health states have low amplitude (gradual slope) with decreasing direction.

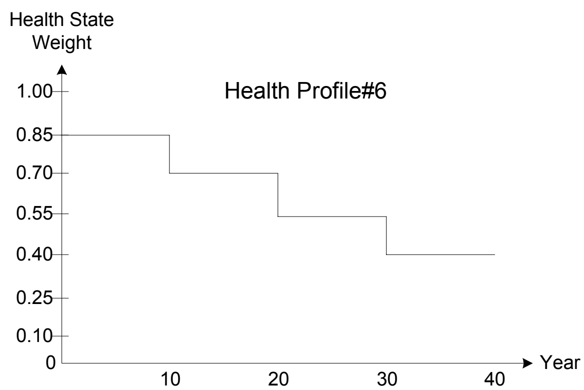


Figure 3.19: Health profile 6 for Phase 2 experiment

Health profile 7 is composed of four health states with 10-year duration for each health state. The changes between each two consecutive health states have low amplitude (gradual slope) with increasing direction.

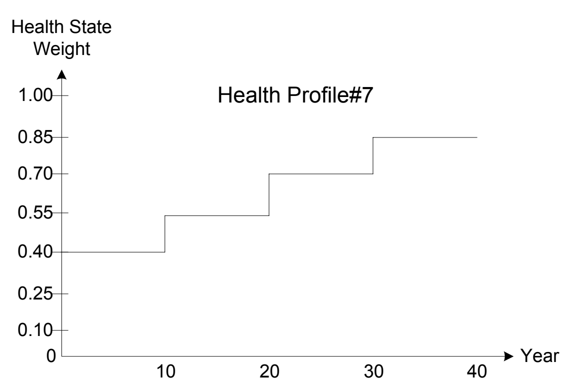


Figure 3.20: Health profile 7 for Phase 2 experiment

Health profile 8 is composed of four health states with 10-year duration for each health state. The changes between each two consecutive health states have high amplitude (steep slope) with decreasing direction.

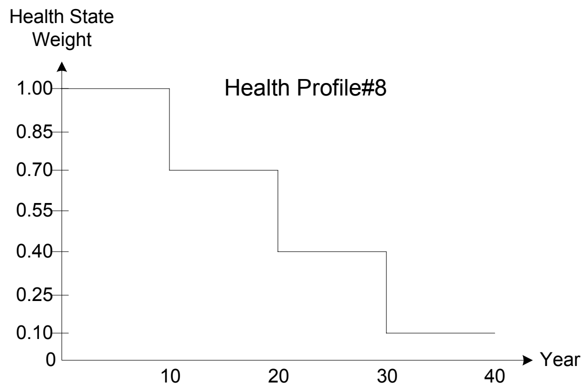


Figure 3.21: Health profile 8 for Phase 2 experiment

Health profile 9 is composed of four health states with 10-year duration for each health state. The changes between each two consecutive health states have high amplitude (steep slope) with increasing direction.

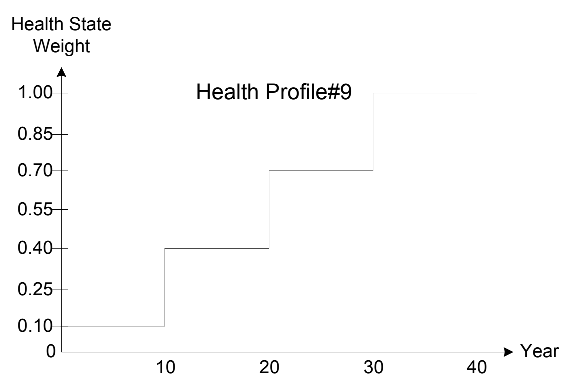


Figure 3.22: Health profile 9 for Phase 2 experiment

Health profile 10 is composed of four health states with 10-year duration for each health state. The profile has no systematic pattern. The changes between each two consecutive health states compose of a loss with high amplitude and gains with both low and high amplitudes.

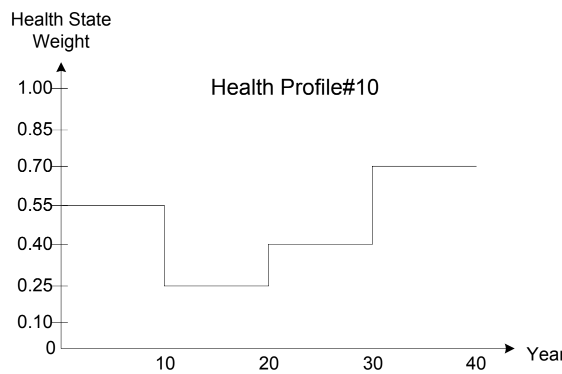


Figure 3.23: Health profile 10 for Phase 2 experiment

The holistic preference scores for the 10 health profiles are assessed by presenting each of the 10 health profiles and asking the subjects to evaluate each health profile. Thus, for each subject, 10 holistic scores for the 10 profiles are obtained. The unconditional preference scores for the 10 health profiles above are calculated from individual assessments of each individual component health state obtained in Phase 1. Then, the duration-weighted scores for each of the 10 health profiles are calculated from Phase 1 data. Regarding the conditional preference scores, most of the conditional preference scores for each pair of two consecutive health states in 10 health profile are obtained from Phase1 experiment. Scenarios that are needed for Phase 2 but are not obtained in Phase 1 are added in the experiment. Table 3.3 lists the additional health scenarios required by each health profile for Phase 2 experiment.

Table 3.3: List of the additional health scenarios required by each health profile for Phase 2 experiment

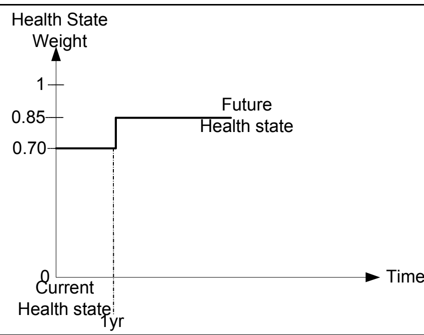
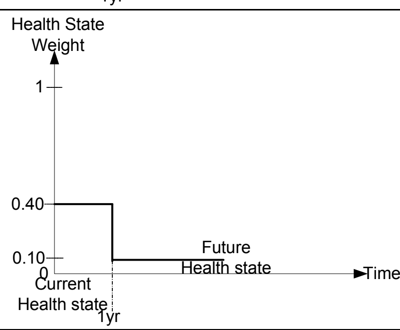
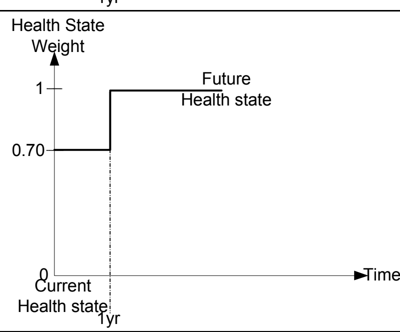
Health Profiles	**Additional Scenarios**
Profile#1	NONE
Profile#2	
Profile#3	
Profile#4	

Table 3.3 (continued): List of the additional health scenarios required by each health profile for Phase 2 experiment

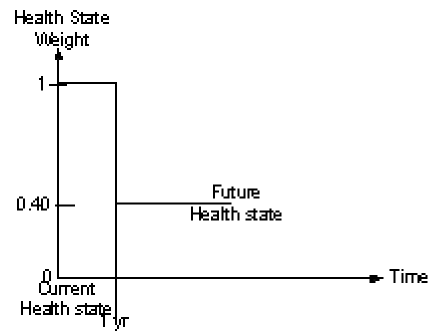
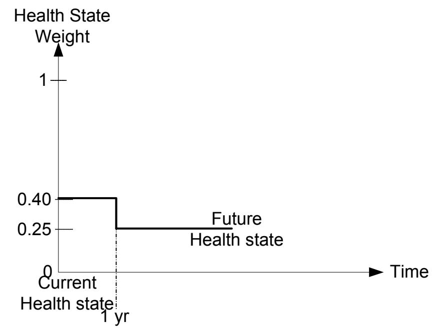
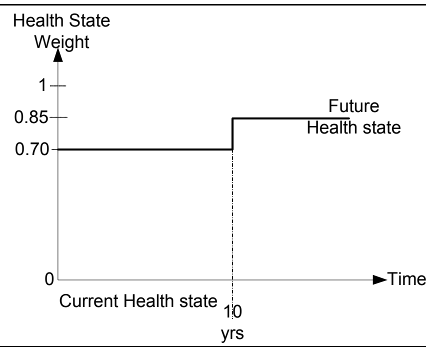
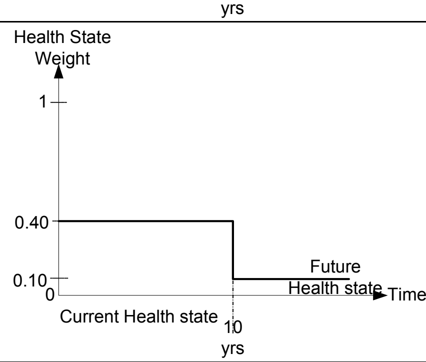
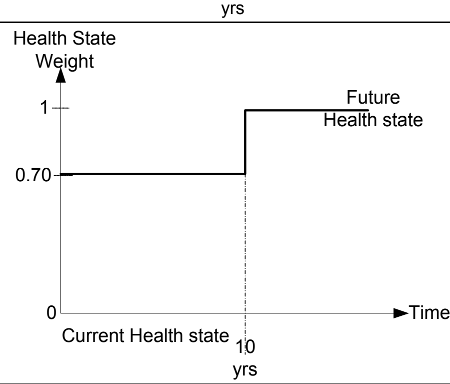
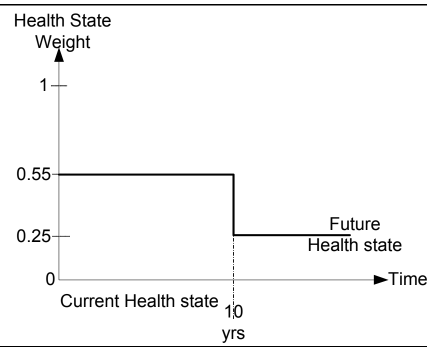
Health Profiles	**Additional Scenarios**
Profile#5	 
Profile#6	NONE
Profile#7	
Profile#8	
Profile#9	

Table 3.3 (continued): List of the additional health scenarios required by each health profile for Phase 2 experiment

Health Profiles	**Additional Scenarios**
Profile#10	

From Table 3.3, there are two health scenarios that could be eliminated. Those are health scenarios that change from level 0.70 (current health state) to level 1.00 (future health state) for the health profile 4 and 9 since the conditional evaluations for these scenarios are for health state at level 1.00 which is perfect health. Thus, the conditional preference scores for these two health scenarios are assumed to be 1.00. Therefore, another 7 health scenarios are added to the experiment.

3.6 Elicitation Technique

The elicitation technique used in the experiment is the visual analog scale (VAS). In the VAS assessment, the subjects are instructed to indicate how they feel about the health states in questions by marking a 100-scale with 2 extremes: best imaginable health state (score 100) and worst imaginable health state (score 0). The VAS is based on a concept that the valuations of health are continuous. It thus can measure the valuations of health state more adequately than categorical scales.

3.7 Development of the Experiment

The experiment starts with designing the experimental scenarios that are presented to the subjects. Two alternative techniques are considered for presenting the scenarios: graphical representation and descriptive representation. The graphical representation uses a graph as a tool to show the health scenarios by using the y-axis to indicate the level of quality of life or health state weight. The descriptive representation uses a short description of the health scenarios with the intention that the subjects can imagine themselves being in those health scenarios while assessing the conditional preference scores. The graphical representation is easy to design and visualize. Especially regarding the amplitude, the amount of amplitude perceived tends to be consistent across subjects when presented in graphs since the levels of health that the subjects perceived are put on the same scale (from 0 to 1 on y-axis). However, when assessing the subjects' preference scores, subjects may tempted to perform calculations based on the position of the health states on the graphs rather than thinking about how they actually feel. Thus, using graphical representation has a high chance that the responses from the subjects will not necessarily represent their actual feelings and preferences. Thus, a descriptive representation may be more appealing. The descriptive representation however is more difficult to design since subjects may value the same health state differently. Selection of health states to be used in the experiment needs to be done in a thoughtful process to ensure that factors to be explored (direction, amplitude, and duration) correspond to the health states selected and can be correctly perceived across all subjects, as designed. For example, while a particular subject perceive that a change from health state A to health state B is of "large" amplitude, as intended by design, a different subject may perceive the same change as "small", and defeats the design.

To ensure that proper descriptive health states are used in the experiment, I decided to use health states as described in the EQ-5D system, a system developed by the EuroQol Group (1990), an international research network established in 1987 by researchers from Finland, the Netherlands, Sweden, and the United Kingdom, for self-health assessment. The EQ-5D presents health state in terms of five dimensions: mobility, self-care, usual activities (work, study, housework, family, or leisure), pain or discomfort, and anxiety or depression. Each dimension is subdivided into three categories, which indicate whether the subject has no problem, a moderate problem, or an extreme problem. For example, health state 21133 indicates that an individual have some problems in walking about, no problems with self-care, no problems with performing usual activities, extreme pain or discomfort, and is extremely anxious or depressed. Valuations of health states in the EQ-5D system have been obtained from several European general population studies including studies of population health status in the United Kingdom (Kind, Dolan, Gudex, and Williams, 1998; Dolan, 1997), the Netherlands (Essink-Bot, Stouthard, and Bonsel, 1993; Busschbach, McDonnell, Essink-Bot, and van Hout, 1999; Busschbach, McDonnell, and van Hout, 1997), Finland (Ohinmaa, Eija, and Sintonen, 1996), and Spain (Rué and Badia, 1996). An example of mean values of 25 health states in EQ-5D system from the EuroQol Rotterdam 1991 survey is shown in Table 3.4 (Busschbach et al., 1999). The values are based on a scale ranging from 0 (worst imaginable health state) to 100 (best imaginable health state).

Table 3.4: Empirical mean values of the EQ-5D health states from the EuroQol Rotterdam 1991 survey (from Busschbach et al., 1999)

EQ-5D Health States	Mean	SD
11111	92.25	13.63
11211	80.47	14.38
11121	73.58	18.48
11112	73.44	18.69
12111	67.93	23.68
21111	62.93	23.18
11221	65.48	18.09
11122	59.95	20.65
21211	52.90	23.51
12212	52.71	20.13
21212	48.50	20.25
32211	45.17	23.33
21232	35.11	23.87
23223	29.89	22.56
22233	27.06	23.15
33321	26.31	23.01
22323	25.96	22.98
32233	24.87	23.44
22333	24.81	22.71
23332	21.24	21.31
32333	20.79	22.65
33332	20.65	21.93
33233	19.80	21.46
23333	15.67	20.80
33333	13.33	23.08

The EQ-5D system is selected for the experiment since the description of health state in the five dimensions is easy to understand and previous studies have been performed to assist in the selection of appropriate health states for this study. Moreover, the subjects are not required to have particular knowledge regarding diseases or any specific disease-related health conditions.

Because different people may have different judgment regarding the same health states, designing the experiment by using the same set of health states across the subjects may result in uncontrollable perceived amplitude, one of the factors that are investigated and thus needs to be controlled. For example, a change from health state A to health state B may be perceived as large amplitude by one subject but as small amplitude by another. Similarly, subjects typically make use of the VAS scale ranging from 0 to 100 in very different ways. For example a difference of 15 points on the scale may be perceived as "large" by some subjects but "small" by others. Hence a within-subject standardization process is used to choose appropriate scenarios for each subject. A set of health states that is used in the experiment has to be selected for each particular subject. In the experiment, each subject thus is presented with his/her customized experimental scenarios.

Therefore, at the beginning of the experiment, each subject is presented with 20 descriptions written on 20 index cards, each representing an individual health state. The subject first ranks-order the cards and then evaluates each of them using VAS. The same set of 20 health states are used for all subjects. Scores are linearly normalized using mapping best health state (perfect health) and worst health state (death) to 1.00 and 0.00 respectively. Then, using normalized scores for each subject, health states that have normalized scores that are closest to the scores needed for the experiment (1.00, 0.85, 0.70, 0.55, 0.40, 0.25, and 0.10) are selected for use in the remainder of the experiment for that particular subject. Thus, while all subjects start with the same 20 states, the final subset of states chosen for the rest of the experiment may differ from subject to subject but corresponds to the health states and profiles demanded by the designs described earlier.

In selecting health states for the experiment, using the selection rule by choosing health states that have the normalized scores that are closest to the scores needed for the experiment (1.00, 0.85, 0.70, 0.55, 0.40, 0.25 and 0.10) may be found inappropriate in some cases. For example, this could be the case when the 20 normalized rating scores are not spreading out over the 0-100 range. Figure 3.24 below helps explain the possible inappropriateness of this selection rule. Each "X" on the left hand sided-scale represents the rating score for each of the 20 health states. "Xs" that are circled belong to health states that will be selected for the experiment at levels 0.85, 0.70, 0.55 and 0.40. Scale on the right hand side shows the reference points that are used as guides to select health states based on the scores needed for the experiment. In this example, none of the health states are eligible to be selected at levels 0.10 and 0.25.

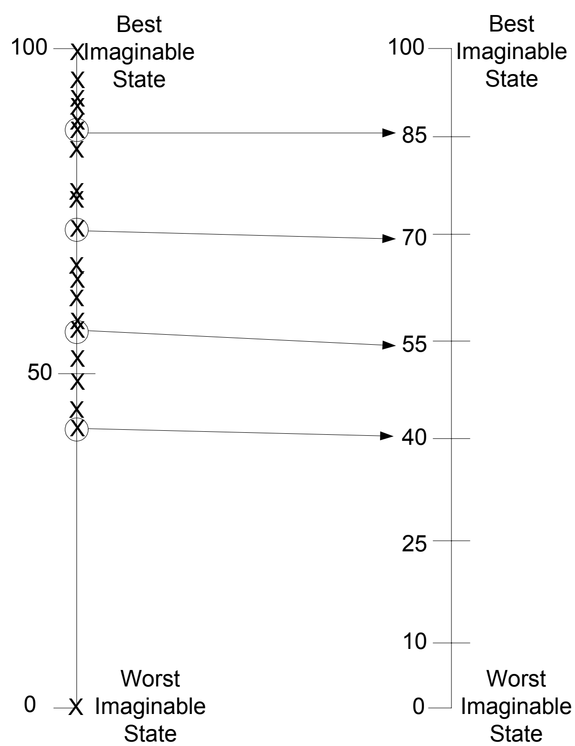


Figure 3.24: An example of a problem of non-scattered rating scores

In order to handle the problem of not sufficiently scattered rating scores, the selection rule is modified as delineated in the following steps:

1. Calculate the range of the 20 normalized rating scores obtained for each subject (the highest score (always be1.00 for perfect health) minus the lowest normalized score that is greater than 0),
2. Divide the range into six portions (since there are six intervals between levels 1.00, 0.85, 0.70, 0.55, 0.40, 0.25 and 0.10), which creates 7 reference points including 1.00.
3. Health states that have the closest scores to the reference points are selected for the experiment.

Table 3.5 shows the formulas to calculate reference points for each health state level required by the design of experiment by using the modified rule.

Table 3.5: Formulas used to calculate reference points using the modified rule

Health State Required Levels	**Reference levels (Modified Rule)**
1.00	1.00
0.85	1.00 – (1.00 – (lowest normalized score > 0))/6
0.70	1.00 – 2*(1.00 – (lowest normalized score > 0))/6
0.55	1.00 – 3*(1.00 – (lowest normalized score > 0))/6
0.40	1.00 – 4*(1.00 – (lowest normalized score > 0))/6
0.25	1.00 – 5*(1.00 – (lowest normalized score > 0))/6
0.10	Lowest normalized score > 0

One may argue that selecting the health states for the experiment by using this modified rule cannot control the amount of amplitude across the subjects. However, what is more important in this study is the relative amplitude that each subject perceives. Each subject uses a different scale but captures “differences” between health states in a different way. By using the modified rule, “large” and “small” amplitude for each subject equals one-third and one-sixth of his/her own scale respectively.

A computer program is developed to use in data collection employs the Visual Basic programming language. All data is recorded in Microsoft Access database. Visual analog scale is represented by a sliding bar on a 0 to 100 scale that allows the subjects to slide the bar to the point that best represents their valuations regarding the health scenarios in questions. Screenshots of the program are attached in the Appendix A. The program is self-explanatory so that the subjects can perform the tasks by themselves.

To summarize, three major parts of the experiment are as follows:

1. Unconditional preference assessments with 20 health states

First, as mentioned above, the subjects are asked to rank order the 20 health states from the best to the worst health states. This part is done manually by using cards such that the subjects are able to manipulate the cards according to their preferences. Then, unconditional preference assessments are obtained for each of the 20 health states. The scores from the assessments are then entered into the computer program for normalization and selection of health states to be used in the later parts. In this first part, basic demographic information is also collected from each subject (gender, ethnicity, age, major field of study, degree worked on, experience with major health issue, general health).

2. Conditional preference assessments with 31 pairs of health states (24 pairs necessary for Phase 1 plus an additional 7 scenarios necessary for profiles introduced in Phase 2)

Based on the 20 unconditional preference scores from the previous assessments, the computer program devised for this study then selects the required health states and generates health scenarios for conditional preference assessments. The 31 health scenarios are presented in a random order. Each scenario displays the hypothetical current health state along with the current health state duration followed by a future health state. The subject is asked to assess the future health state in the scenario given the current scenario and rate it using VAS.

3. Holistic preference assessments with 10 health profiles

The 10 health profiles are generated by the computer program based on the 20 unconditional preference scores in the first part. The 10 health profiles are divided into two groups based on the profiles duration: five 4-year profiles, and five 40-year profiles. The scenarios of the 4-year profiles are framed in terms of the conditions during an illness lasting 4 years, while the 40-year profiles are framed as lifetime profiles. Before the holistic preference assessment starts, the subjects perform paired-comparisons for all possible combinations of the five profiles within each group with the aim of generating the ranking among the five profiles. Then, the subjects are asked to assess each health profile.

3.8 Study Subjects

The subjects in this study are undergraduate or graduate students of the School of Industrial and Systems Engineering at the Georgia Institute of Technology. This subject pool represents a young and relatively healthy population. One rationale behind using a healthy population in the study is that in economic evaluation of health care, the valuations of health used in cost effectiveness analysis are societal and thus include healthy subjects as opposed to only patients with actual experience with diseased or morbid states. The values of health determined by a general population are more valid than those determined by patients in the context of decisions on the alternative allocation of resources (Hadorn, 1991). While the sample to be used in this study is certainly not representative of the general population, it represents one segment of the general population. More importantly, the purpose of this exploratory study is to investigate, develop, and test a new assessment approach, not to generate utility assessments that can be used in actual cost-effectiveness studies. Thus a student population is deemed appropriate for this study.

The sample size needed in the study is determined by power analysis. Three values that need to be determined in order to estimate the sample size are power value (the probability of correctly rejecting the null hypothesis when it is false), minimum effect size, and standard deviation. Table 3.6 below shows the estimated sample size at each power level (0.80, 0.85, 0.90, 0.95) and at each minimum effect size (0.025, 0.050, 0.075, and 0.100), calculated at standard deviation equals 0.20 with alpha at 0.05. The standard deviation was estimated from the descriptive statistics of raw VAS scores in the study by Busschbach et al. (1999) as shown previously in Table 3.4. Also, Figure 3.25

graphs the estimated sample size for each power level and effect size corresponding to Table 3.6. From Table 3.6 and Figure 3.25, it appears that a sample size of 70 gives at least 80% probability of detecting an effect size as small as 0.05. A sample size of 80 to 100 is therefore targeted for this study.

Table 3.6: Estimated sample size calculated at alpha = 0.05 and standard deviation = 0.20 for each power level and minimum effect size

Power level	**Effect size**	**Estimated Sample Size**
0.95	0.025	417
	0.050	105
	0.075	47
	0.100	27
0.90	0.025	337
	0.050	85
	0.075	38
	0.100	22
0.85	0.025	288
	0.050	73
	0.075	33
	0.100	19
0.80	0.025	252
	0.050	64
	0.075	29
	0.100	16

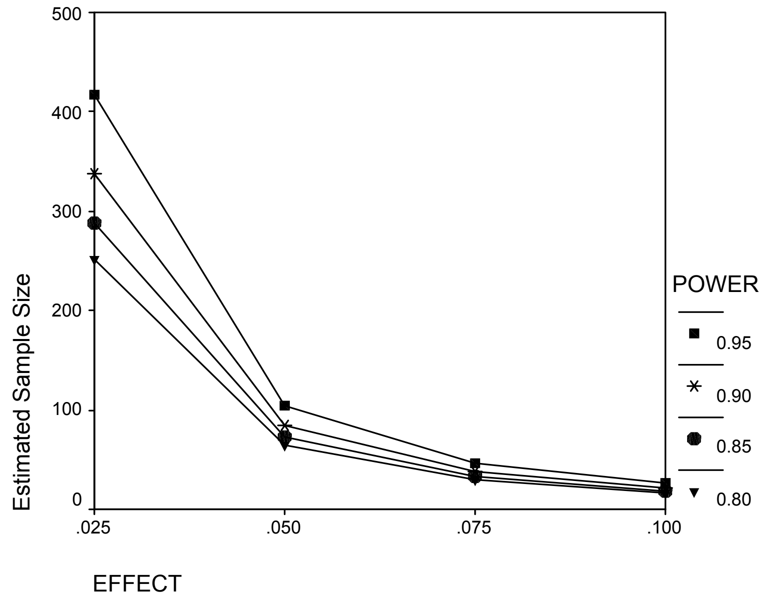


Figure 3.25: Graph represents estimated sample size calculated at alpha = 0.05 and standard deviation = 0.20

3.9 Study Protocol

The study is advertised through electronic mail by targeting undergraduate and graduate students of the School of Industrial and Systems Engineering at the Georgia Institute of Technology. Those interested in participating in the study make an appointment with the investigator of the study and come to perform the study at the Laboratory for Human-Computer Interaction and Health Care Informatics in the School of Industrial and Systems Engineering. The experiment procedure starts by giving a short explanation of the study and asking if the student agrees in participating in the study. A signed consent form approved by the Georgia Institute of Technology's Institutional Review Board is obtained for each participant. Each subject sits at a personal computer station in the Laboratory for Human-Computer Interaction and Health Care Informatics and completes the tasks. Each subject who completes the entire experiment procedure receives $20 as compensation.

CHAPTER 4

ANALYSIS AND RESULTS

4.1 Data Description

Data were collected from 100 undergraduate and graduate students mainly from the School of Industrial and Systems Engineering at the Georgia Institute of Technology. Data for each subject consisted of the followings:

- Background information (gender, ethnicity, age, major field of study, degree worked on, experience with major health issue, and general health)
- 20 unconditional preference scores for the 20 EQ-5D health states
- 31 conditional preference scores for 31 pairs of health states (24 pairs necessary for Phase 1 experiment and an additional 7 pairs for health profiles introduced in Phase 2)
- 10 holistic preference scores for 10 health profiles in Phase 2 experiment
- 10 rankings from paired-comparisons for the 10 health profiles in Phase 2

Data from one of the 100 subjects was removed prior to the analysis due to an experimental error, leaving a potential 99 usable cases.

4.2 Analysis and Results for Phase 1 Experiment

Phase 1 experiment involves exploring the relationship between two consecutive health states, determining whether a preference score for a future health state is dependent upon the level of the current health state and if so, what factors affect the preference and in what ways. Factors of interest that are included in the experiment are duration of current health state (2 levels: 1 year and 10 years), direction of change from current health state (2 levels: decrease and increase), and amplitude of change from current to future health state (2 levels: small and large). Three replications of the 2^3 factorial design at three different levels of future health states (high (level 0.70), medium (level 0.55), and low (level 0.40)) are performed. Those three sets of the 2^3 factorial design are referred to as Design A, B, and C, respectively.

4.2.1 Data Preparation

Data used in the analysis for Phase 1 consist of the 24 conditional preference scores obtained per subject from the three sets of the 2^3 factorial design at three different levels of future health states. To compare scores across subjects, conditional preference scores were linearly normalized for each subject by mapping best health state (perfect health) and worst health state (death), specified in the stage of unconditional preference assessments, to 1.00 and 0.00 respectively. Seven cases, for which normalized scores were higher than 1.00 or lower than 0.00, were removed. After normalization, outlier analysis was performed using box plot. Responses identified as extreme outliers (values exceed three box lengths (interquartile range) from the lower and upper edge of the box) were removed. Box plots and descriptive statistics of the subjects whose data were removed for Phase 1 analysis are provided in Appendix B. As a result, data used in Phase 1 analysis correspond to 92 subjects.

4.2.2 Data Description

Background Variables

Phase 1 analysis uses a total of 92 subjects, 36 (39%) female students and 56 (61%) male students. Age ranges from 18 to 33 years, with an average of 22.36 years. Among the 92 students, 44 (48%) are white, 28 (30%) are Asian or Pacific Islander, 10 (11%) are Hispanic, and 6 (7%) are black. The majority (83 (90%)) of them major in Industrial Engineering and 72 (78%) are undergraduate students. Concerning major health issues, 13 (14%) experienced a major health issue themselves, 48 (52%) experienced one in their families, and 40 (43%) experienced one in someone else close to them. Regarding the students health status, none of them indicates having poor health. A summary of the background variables of the 92 subjects is shown in Table 4.1. Figure 4.1 shows the distribution of age, the only continuous variable.

Table 4.1: Summary of Background Variables for Phase 1 Analysis

Variable	**Counts**	**Percent**	**Mean**	**S.D.**	**Range**
Gender					
Female	36	39.13			
Male	56	60.87			
Ethnic Background					
White, not of Hispanic origin	44	47.83			
Black, not of Hispanic origin	6	6.52			
Hispanic	10	10.87			
Asian or Pacific Islander	28	30.43			
Others	4	4.35			
Age (years)			22.36	2.51	18.00 - 33.00
Major field of study					
Industrial Engineering	83	90.22			
Other – non-engineering	6	6.52			
Other – Engineering	3	3.26			
Degree of study					
B.S.	72	78.26			
M.S.	8	8.70			
Ph.D.	11	11.96			
Other	1	1.08			
Health problem in yourself					
Yes, major issue(s)	13	14.13			
Yes, minor issue(s)	42	45.65			
No	37	40.22			
Health problem in your family					
Yes, major issue(s)	48	52.17			
Yes, minor issue(s)	30	32.61			
No	14	15.22			
Health problem in someone else close to you					
Yes, major issue(s)	40	43.48			
Yes, minor issue(s)	25	27.17			
No	27	29.35			
General health condition					
Excellent	26	28.26			
Very good	49	53.26			
Good	16	17.39			
Fair	1	1.09			
Poor	0	0.00			

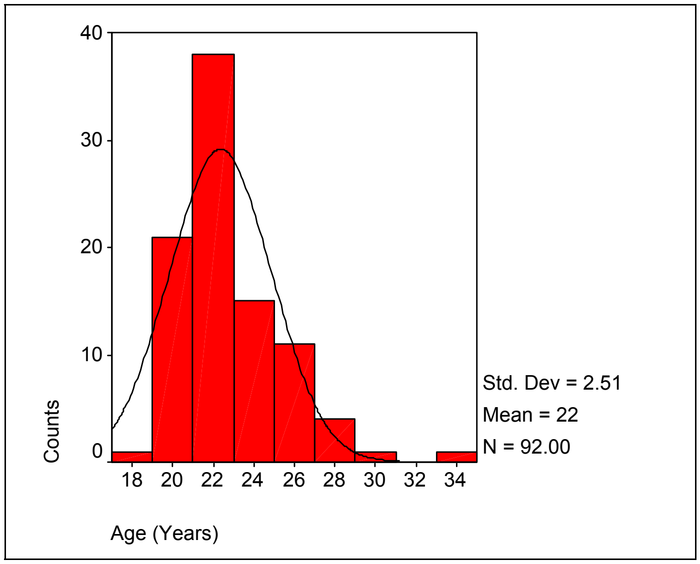


Figure 4.1: Age distribution of subjects in Phase 1 analysis

Unconditional Preference Score

At the beginning of the experiment, each subject was presented with 20 descriptions on 20 index cards, each representing an individual health state in EQ-5D system, as shown in Table 4.2. The subject first rank-ordered the cards and then rated each of them using a visual analog scale (VAS). Thus, each subject generated 20 unconditional preference scores.

Table 4.2: Descriptions of the 20 health states used in the experiment

Health State#	EQ-5D Coding	Description
1	11111	No problems in walking about. No problems with self-care. No problems with performing your usual activities (e.g.work, study, leisure activities, family, housework). No pain or discomfort. Not anxious or depressed.

Table 4.2 (continued): Descriptions of the 20 health states used in the experiment

Health State#	**EQ-5D Coding**	**Description**
2	11211	No problems in walking about. No problems with self-care. Some problems with performing your usual activities (e.g.work, study, leisure activities, family, housework). No pain or discomfort. Not anxious or depressed.
3	11121	No problems in walking about. No problems with self-care. No problems with performing your usual activities (e.g.work, study, leisure activities, family, housework). Moderate pain or discomfort. Not anxious or depressed.
4	12111	No problems in walking about. Some problems with self-care. No problems with performing your usual activities (e.g.work, study, leisure activities, family, housework). No pain or discomfort. Not anxious or depressed.
5	11221	No problems in walking about. No problems with self-care. Some problems with performing your usual activities (e.g.work, study, leisure activities, family, housework). Moderate pain or discomfort. Not anxious or depressed.
6	11122	No problems in walking about. No problems with self-care. No problems with performing your usual activities (e.g.work, study, leisure activities, family, housework). Moderate pain or discomfort. Moderately anxious or depressed.
7	12222	No problems in walking about. Some problems with self-care. Some problems with performing your usual activities (e.g.work, study, leisure activities, family, housework). Moderate pain or discomfort. Moderately anxious or depressed.

Table 4.2 (continued): Descriptions of the 20 health states used in the experiment

Health State#	EQ-5D Coding	Description
8	12212	No problems in walking about. Some problems with self-care. Some problems with performing your usual activities (e.g.work, study, leisure activities, family, housework). No pain or discomfort. Moderately anxious or depressed.
9	22222	Some problems in walking about. Some problems with self-care. Some problems with performing your usual activities (e.g.work, study, leisure activities, family, housework). Moderate pain or discomfort. Moderately anxious or depressed.
10	13212	No problems in walking about. Unable to wash and dress yourself. Some problems with performing your usual activities (e.g.work, study, leisure activities, family, housework). No pain or discomfort. Moderately anxious or depressed.
11	13311	No problems in walking about. Unable to wash and dress yourself. Unable to perform your usual activities (e.g.work, study, leisure activities, family, housework). No pain or discomfort. Not anxious or depressed.
12	12223	No problems in walking about. Some problems with self-care. Some problems with performing your usual activities (e.g.work, study, leisure activities, family, housework). Moderate pain or discomfort. Extremely anxious or depressed.
13	21232	Some problems in walking about. No problems with self-care. Some problems with performing your usual activities (e.g.work, study, leisure activities, family, housework). Extreme pain or discomfort. Moderately anxious or depressed.

Table 4.2 (continued): Descriptions of the 20 health states used in the experiment

Health State#	EQ-5D Coding	Description
14	23223	Some problems in walking about. Unable to wash and dress yourself. Some problems with performing your usual activities (e.g.work, study, leisure activities, family, housework). Moderate pain or discomfort. Extremely anxious or depressed.
15	22233	Some problems in walking about. Some problems with self-care. Some problems with performing your usual activities (e.g.work, study, leisure activities, family, housework). Extreme pain or discomfort. Extremely anxious or depressed.
16	23332	Some problems in walking about. Unable to wash and dress yourself. Unable to perform your usual activities (e.g.work, study, leisure activities, family, housework). Extreme pain or discomfort. Moderately anxious or depressed.
17	33332	Confined to bed. Unable to wash and dress yourself. Unable to perform your usual activities (e.g.work, study, leisure activities, family, housework). Extreme pain or discomfort. Moderately anxious or depressed.
18	23333	Some problems in walking about. Unable to wash and dress yourself. Unable to perform your usual activities (e.g.work, study, leisure activities, family, housework). Extreme pain or discomfort. Extremely anxious or depressed.
19	33333	Confined to bed. Unable to wash and dress yourself. Unable to perform your usual activities (e.g.work, study, leisure activities, family, housework). Extreme pain or discomfort. Extremely anxious or depressed.
20	Death	Death

After obtaining the 20 unconditional preference scores, scores were linearly normalized for each subject by mapping best health state (health state #1 - perfect health) and worst health state (health state #20 - death) to 1.00 and 0.00 respectively. Descriptive statistics of the normalized unconditional preference score for each health state from 92 subjects are summarized in Table 4.3. Means of normalized unconditional preference scores of the 20 health states are distributed approximately evenly over the scale as can be seen in Figure 4.2. From Table 4.3, one can notice that the maximum standard deviation obtained is 0.18, which is less than the standard deviation of 0.20 which was used in determining power and sample size for this experiment at the earlier stage. In recalculating power for this experiment with a sample size of 92 and a standard deviation of 0.18, this experiment provides at least 96.43% probability of detecting an effect size as small as 0.05.

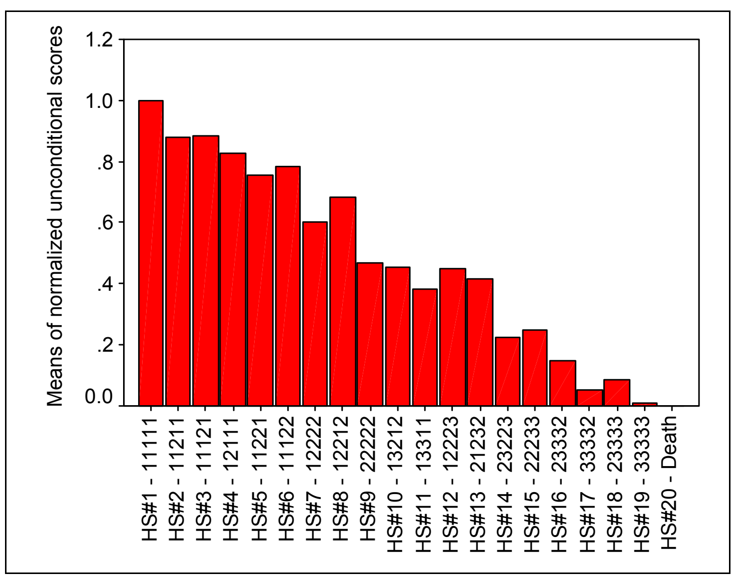


Figure 4.2: Means of normalized unconditional preference scores for the 20 health states

Table 4.3: Descriptive statistics for normalized unconditional preference scores

Health State	N	Min	Max	Mean	S.D.
Health State#1 - 11111 (perfect health)	92	1.00	1.00	1.00	0.00
Health State#2 – 11211	92	0.38	0.98	0.88	0.09
Health State#3 – 11121	92	0.65	0.99	0.89	0.08
Health State#4 – 12111	92	0.50	0.98	0.83	0.10
Health State#5 – 11221	92	0.33	0.99	0.76	0.11
Health State#6 – 11122	92	0.38	0.98	0.78	0.11
Health State#7 – 12222	92	0.29	0.84	0.60	0.10
Health State#8 – 12212	92	0.30	0.91	0.69	0.10
Health State#9 – 22222	92	-0.40	0.77	0.47	0.15
Health State#10 – 13212	92	0.09	0.80	0.45	0.15
Health State#11 – 13311	92	-0.13	0.76	0.38	0.18
Health State#12 – 12223	92	0.13	0.74	0.45	0.14
Health State#13 – 21232	92	-0.20	0.76	0.42	0.18
Health State#14 – 23223	92	-0.60	0.53	0.23	0.14
Health State#15 – 22233	92	-0.60	0.70	0.25	0.17
Health State#16 – 23332	92	-0.80	0.50	0.15	0.15
Health State#17 – 33332	92	-1.00	0.30	0.05	0.14
Health State#18 – 23333	92	-0.80	0.30	0.09	0.13
Health State#19 – 33333	92	-1.00	0.11	0.01	0.12
Health State#20 – Death	92	0.00	0.00	0.00	0.00

Based on the 20 normalized unconditional preference scores, the computer program created for this study then selected the required health states in order to generate health scenarios for conditional and holistic preference assessments. For each subject, 7 out of 20 health states were selected. While all subjects started with the same 20 health states, the subset of health states chosen for the rest of the experiment differed from subject to subject but corresponded to the health states required by the design of the experiment explained in Chapter 3. The required health states were for levels 1.00, 0.85, 0.70, 0.55, 0.40, 0.25, and 0.10. Table 4.4 shows the frequencies at which each health state was selected as one of the required health states. By design, health state #1 (perfect health) was used as the level 1.00 health state for all subjects. However, for health state level 0.85, health states #2 (11211) and 3 (11121) while most frequently chosen (23% of the time for each health state) were not *always* picked. For health state level 0.70, health state #8 (12212) was the most frequently used health state (26%). Health state #9 (22222) was the most frequent selection (22%) for health state level 0.55. Health state #12 (12223) was selected 20% of the time to represent health state level 0.40. Finally, for health state level 0.25, and 0.10, health states #14 (23223) and 19 (33333) were the most frequently selected health states with the frequencies of 25% and 64%, respectively.

Table 4.4: Frequencies of health states that are selected to generate health scenarios in the experiment

Health State#	**Level 1.00**		**Level 0.85**		**Level 0.70**		**Level 0.55**		**Level 0.40**		**Level 0.25**		**Level 0.10**	
	Freq.	**%**	**Freq.**	**%**	**Freq.**	**%**	**Freq.**	**%**	**Freq.**	**%**	**Freq.**	**%**	**Freq.**	**%**
1	92	100												
2			21	23	2	2	2	2						
3			21	23	1	1								
4			14	15	5	5			1	1				
5			16	17	16	17	2	2						
6			14	15	9	10			1	1				
7			1	1	20	22	13	14	5	5				
8			5	5	24	26	5	5						
9					2	2	20	22	9	10	1	1		
10					1	1	13	14	14	15	6	7		
11					5	5	10	11	9	10	8	9	1	1
12					4	4	12	13	18	20	4	4		
13					2	2	14	15	10	11	7	8	1	1
14									12	13	23	25	5	5
15					1	1	1	1	9	10	15	16	5	5
16									4	4	18	20	4	4
17											2	2	7	8
18											8	9	10	11
19													59	64
20														

Additionally, Table 4.5 summarizes the descriptive statistics for unconditional preference scores for health states that were selected for the experiment across all the 92 subjects. It can be seen that the average scores of the health states that were selected are very close to the scores required by the design of experiment.

Table 4.5: Unconditional preference score for each level of health state selected for the experiment

Health State	N	Min	Max	Mean	S.D.
Health State Level 1.00	92	1.000	1.000	1.000	0.000
Health State Level 0.85	92	0.750	0.904	0.842	0.029
Health State Level 0.70	92	0.610	0.785	0.686	0.032
Health State Level 0.55	92	0.375	0.620	0.522	0.035
Health State Level 0.40	92	0.268	0.522	0.365	0.047
Health State Level 0.25	92	0.100	0.300	0.208	0.042
Health State Level 0.10	92	0.010	0.200	0.057	0.036

Conditional Preference Score (CPS)

Conditional preference scores were obtained by presenting each subject with a set of health scenarios. A health scenario consisted of a description of a current health state along with the duration that the subject was asked to imagine himself/herself currently experiencing, followed by a description of a future health state that the subject would experience starting the next day. Each subject was asked to give a CPS for the future health state on a visual analog scale (VAS).

Table 4.7, Table 4.9, and Table 4.11 show descriptive statistics of CPS obtained from all subjects for Design A, B, and C for all scenarios (8 scenarios per design), along with the results from Kolmogorov-Smirnov normality tests. Detailed scenarios of Design A, B, and C are shown in Table 4.6, Table 4.8, and Table 4.10, respectively. Means of

CPS are ranging from 0.410 to 0.674 for Design A, from 0.480 to 0.545 for Design B, and from 0.394 to 0.442 for Design C. The results from Kolmogorov-Smirnov normality tests show that all but two scenarios in Design A are normally distributed. However, one of the two scenarios that are not normally distributed is close to being normal (scenario 3-A, p-value = 0.049). Another non-normally distributed scenario is scenario 4-A (p-value = 0.011). Figure 4.3, Figure 4.4, and Figure 4.5 illustrate the distributions of CPS for Design A, B, and C, for each scenario, respectively.

Table 4.6: Design A scenarios

Design A Scenarios	**Current Health State Duration**	**Current Health State Level**	**Future Health State Level**
1-A	1 year	1.00	0.70
2-A	10 years	1.00	0.70
3-A	1 year	0.85	0.70
4-A	10 years	0.85	0.70
5-A	1 year	0.40	0.70
6-A	10 years	0.40	0.70
7-A	1 year	0.55	0.70
8-A	10 years	0.55	0.70

Table 4.7: Conditional Preference Score for Each Scenario in Design A

Design A Scenarios (Future Health State Level = 0.70)	N	Min	Max	Mean	S.D.	Test of Normality (Kolmogorov-Smirnov)	
						Statistics	p-value
1-A	92	0.021	0.857	0.410	0.179	0.075	0.200
2-A	92	0.225	0.958	0.663	0.168	0.066	0.200
3-A	92	0.288	0.919	0.674	0.121	0.093	0.049
4-A	92	0.256	0.914	0.670	0.137	0.107	0.011
5-A	92	0.298	0.985	0.645	0.147	0.046	0.200
6-A	92	0.284	1.000	0.644	0.162	0.043	0.200
7-A	92	0.267	0.929	0.644	0.127	0.055	0.200
8-A	92	0.328	0.983	0.644	0.144	0.061	0.200

Table 4.8: Design B scenarios

Design B Scenarios	Current Health State Duration	Current Health State Level	Future Health State Level
1-B	1 year	0.85	0.55
2-B	10 years	0.85	0.55
3-B	1 year	0.70	0.55
4-B	10 years	0.70	0.55
5-B	1 year	0.25	0.55
6-B	10 years	0.25	0.55
7-B	1 year	0.40	0.55
8-B	10 years	0.40	0.55

Table 4.9: Conditional Preference Score for Each Scenario in Design B

Design B Scenarios* (Future Health State Level = 0.55)	N	Min	Max	Mean	S.D.	Test of Normality (Kolmogorov-Smirnov)	
						Statistics	p-value
1-B	92	0.099	0.840	0.545	0.130	0.068	0.200
2-B	92	0.147	0.870	0.539	0.138	0.080	0.185
3-B	92	0.270	0.910	0.535	0.128	0.077	0.200
4-B	92	0.181	0.868	0.534	0.131	0.080	0.188
5-B	92	0.122	0.924	0.480	0.157	0.056	0.200
6-B	92	0.152	0.948	0.518	0.160	0.071	0.200
7-B	92	0.280	0.910	0.534	0.122	0.057	0.200
8-B	92	0.282	0.914	0.538	0.130	0.091	0.056

Table 4.10: Design C scenarios

Design C Scenarios	Current Health State Duration	Current Health State Level	Future Health State Level
1-C	1 year	0.70	0.40
2-C	10 years	0.70	0.40
3-C	1 year	0.55	0.40
4-C	10 years	0.55	0.40
5-C	1 year	0.10	0.40
6-C	10 years	0.10	0.40
7-C	1 year	0.25	0.40
8-C	10 years	0.25	0.40

Table 4.11: Conditional Preference Score for Each Scenario in Design C

Design C Scenarios* (Future Health State Level = 0.40)	N	Min	Max	Mean	S.D.	Test of Normality (Kolmogorov-Smirnov)	
						Statistics	p-value
1-C	92	0.133	0.713	0.404	0.133	0.069	0.200
2-C	92	0.113	0.809	0.422	0.142	0.041	0.200
3-C	92	0.050	0.815	0.412	0.124	0.077	0.200
4-C	92	0.042	0.770	0.414	0.136	0.084	0.115
5-C	92	0.092	0.935	0.442	0.210	0.065	0.200
6-C	92	0.069	0.983	0.492	0.206	0.092	0.054
7-C	92	0.130	0.857	0.394	0.144	0.081	0.176
8-C	92	0.074	0.807	0.408	0.153	0.071	0.200

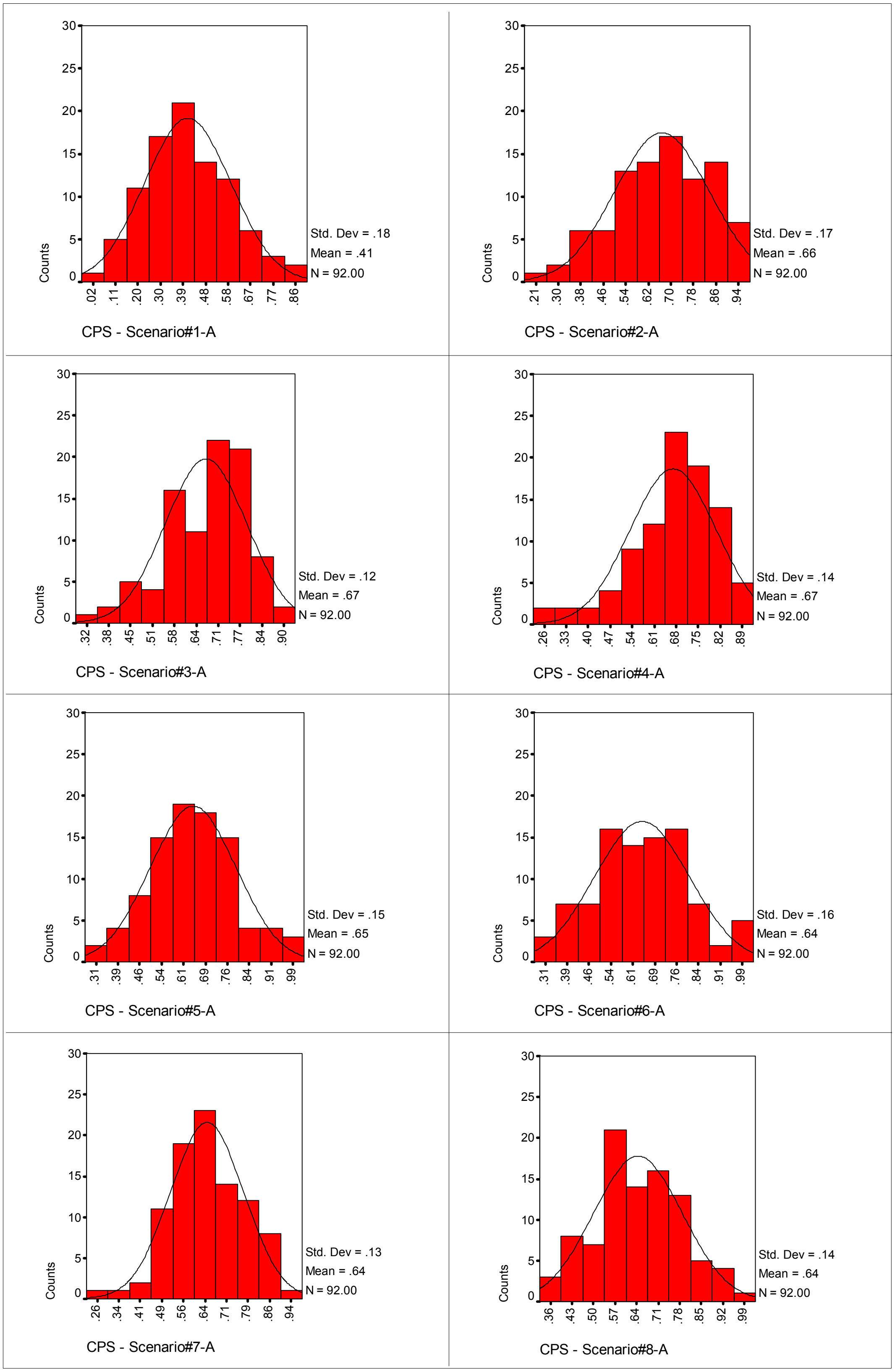


Figure 4.3: Distribution of conditional preference score in each scenario in Design A

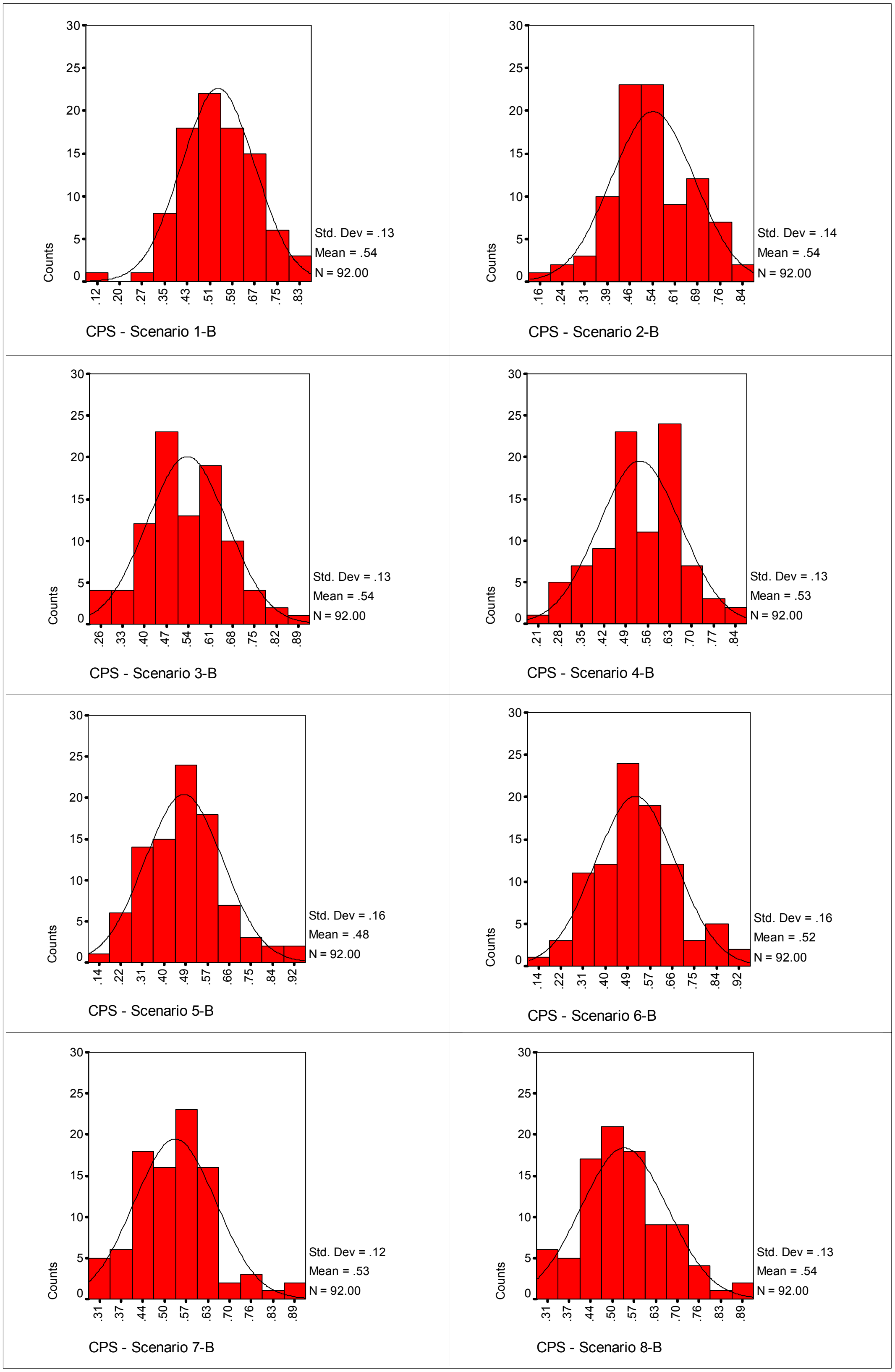


Figure 4.4: Distribution of conditional preference score in each scenario in Design B

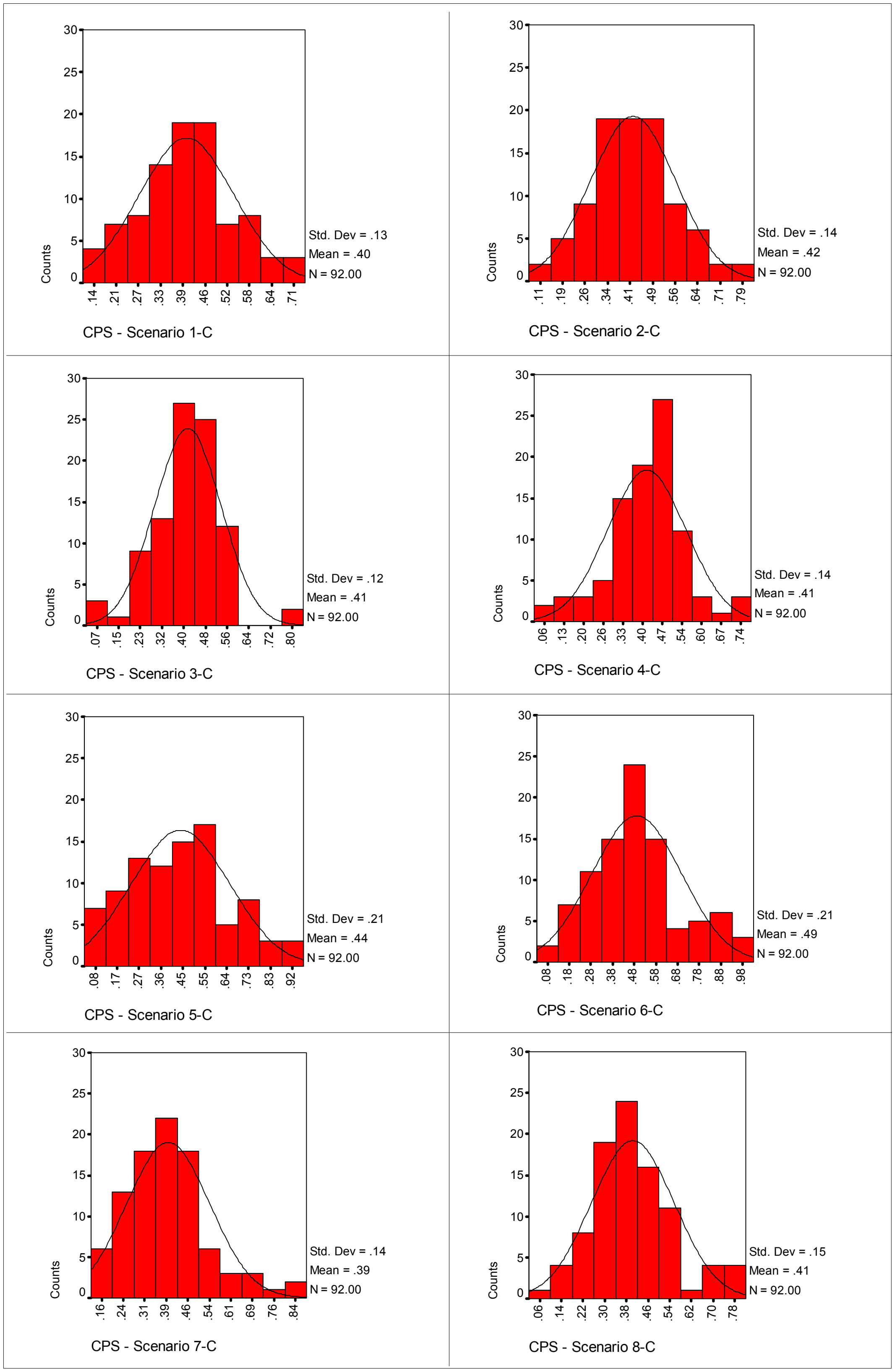


Figure 4.5: Distribution of conditional preference score in each scenario in Design C

4.2.3 Data Analysis

General Linear Model – Repeated Measures (GLM-RM) was used for data analysis in Phase 1. This technique was used as opposed to between-groups analysis of variance (ANOVA) since the same subjects participated in all conditions of the experiment. ANOVA depends on the assumption that scores in different conditions are independent. However, when the same subjects participate in all conditions, this assumption is violated. GLM-RM, however, does not require this assumption but instead, it requires that the level of dependence between experimental conditions has to be roughly equal. This assumption is called sphericity. The sphericity assumption was checked throughout the analysis. SPSS 11.5 for Windows was used as the analysis tool for running GLM-RM.

GLM-RM was performed by treating duration, direction, and amplitude as repeated factors, while conditional preference score is a dependent variable. Three sets of GLM-RM were executed separately for each design (Design A, B, and C). Summary of the main effects, 2-way interaction, and 3-way interaction effects are summarized in Table 4.12. This approach allows testing the hypotheses presented in Chapter 3. In the following sections, results are presented for each hypothesis.

Table 4.12: Results from GLM-RM for each design

	Design A		**Design B**		**Design C**	
	Statistics	**p-value**	**Statistics**	**p-value**	**Statistics**	**p-value**
Duration	F = 47.24	<0.001	F = 1.34	0.250	F = 6.42	0.013
Direction	F = 8.11	0.005	F = 2.87	0.094	F = 2.65	0.107
Amplitude	F = 58.31	<0.001	F = 2.96	0.089	F = 14.48	<0.001
Duration x Direction	F = 62.04	<0.001	F = 3.04	0.085	F = 1.58	0.212
Duration x Amplitude	F = 75.38	<0.001	F = 1.23	0.271	F = 2.72	0.102
Direction x Amplitude	F = 46.91	<0.001	F = 4.21	0.043	F = 9.47	0.003
Duration x Direction x Amplitude	F = 68.63	<0.001	F = 2.06	0.155	F = 0.43	0.511

*The grey highlight indicates significant effect at alpha = 0.05

Hypothesis 1: **There is no significant three-way interaction effect between direction, amplitude and duration on the conditional preference score of the future health state.**

Table 4.13 shows the summary of mean of CPS at each combination of the three factors. The results from GLM-RM (see Table 4.14) show that the three-way interaction effect between direction, amplitude and duration on the CPS of the future health state is not significant at medium (level 0.55, Design B) and low levels (level 0.40, Design C) of future health state. At the high level (level 0.70, Design A) the results reveal a significant three-way interaction effect with p-value less than 0.001. In order to explain the significant three-way interaction effect of Design A, Figure 4.6, Figure 4.7, and Figure 4.8 depict means of CPS from Design A for each combination of the three factors.

Table 4.13: Mean of CPS for each combination of direction, amplitude, and duration, for each design

Direction	Amplitude	Duration	Mean (Design A)	Mean (Design B)	Mean (Design C)
Decrease	Small	1 Year	0.674	0.535	0.412
		10 Years	0.670	0.534	0.414
	Large	1 Year	0.410	0.545	0.404
		10 Years	0.663	0.539	0.422
Increase	Small	1 Year	0.644	0.534	0.394
		10 Years	0.644	0.538	0.408
	Large	1 Year	0.645	0.480	0.442
		10 Years	0.644	0.518	0.492

Table 4.14: Summary of the results from GLM-RM for 3-way interaction effect between duration, direction, and amplitude

3-Way Interaction Effect (duration x direction x amplitude)	**GLM-RM**	
	Test statistic	p-value
Design A (Future health state level = 0.70)	F = 68.63	<0.001
Design B (Future health state level = 0.55)	F = 2.06	0.155
Design C (Future health state level = 0.40)	F = 0.43	0.511

*The grey highlight indicates significant effect at alpha = 0.05

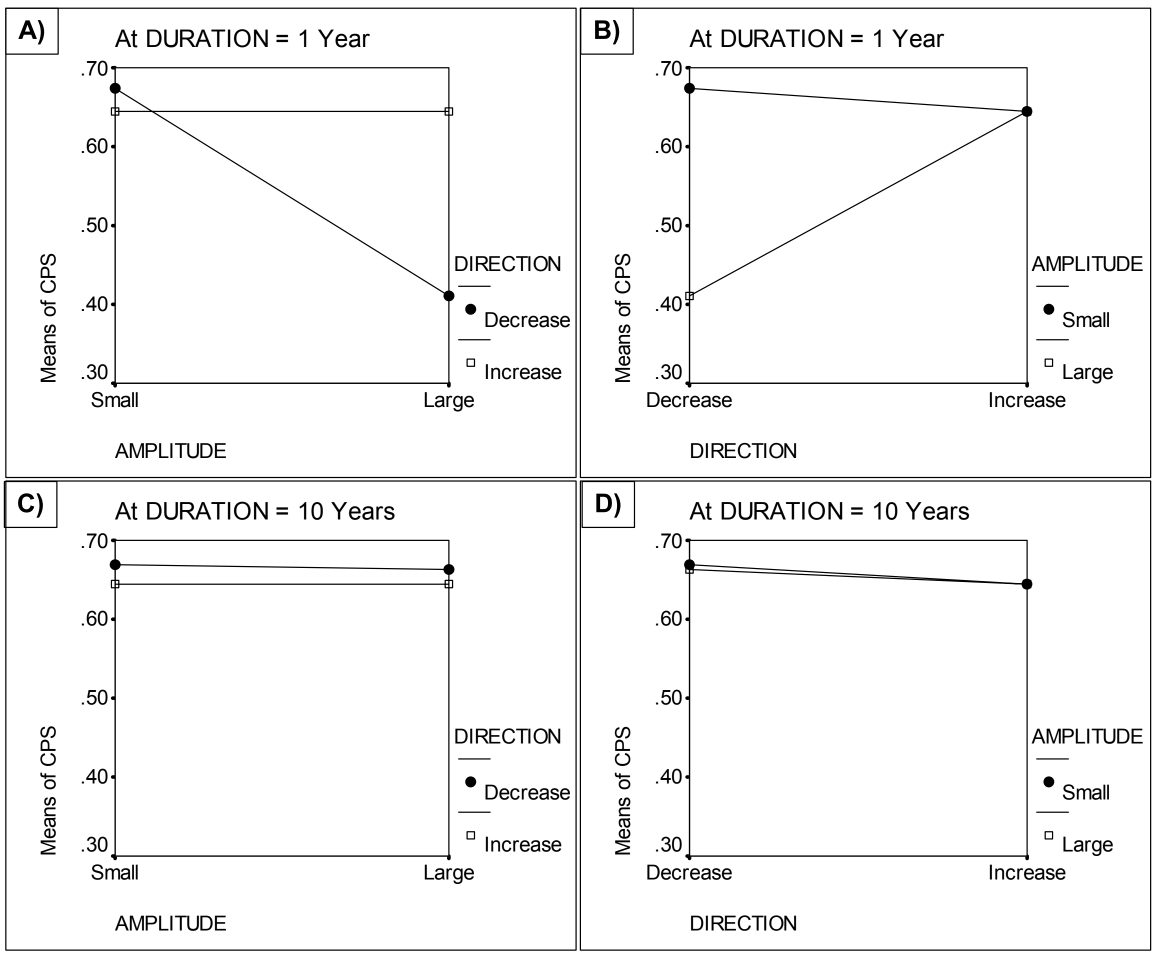


Figure 4.6: Graphs represent three-way interaction effect on CPS for Design A (future health state level = 0.70) for each level of current health state duration

By examining the three-way interaction by level of current health state duration when future health state is at level 0.70 as shown in Figure 4.6, it can be seen that direction and amplitude effects interact only when current health state has a 1-year duration (Figure 4.6A and Figure 4.6B as opposed to Figure 4.6C and Figure 4.6D). Consider Figure 4.6A, when current health state has 1-year duration and future health state has small amplitude of change from current health state, the mean of CPS for a decreased future health state is slightly higher than the mean CPS for an increased future health state (0.674 versus 0.644). However, when current health state has 1-year duration and future health state has large amplitude of change from current health state, changing the direction of change from *decreasing* to *increasing* significantly increases CPS by an average of 0.235.

Considering another aspect (Figure 4.6B), when future health state is decreased from current health state that has 1-year duration, varying amplitude of change from small to large significantly decreases CPS by an average of 0.264. However, when future health state is increased from current health state that has 1-year duration, CPS for future health state that has small amplitude of change and CPS for current health state that has large amplitude of change are approximately equal (0.644 and 0.645, respectively).

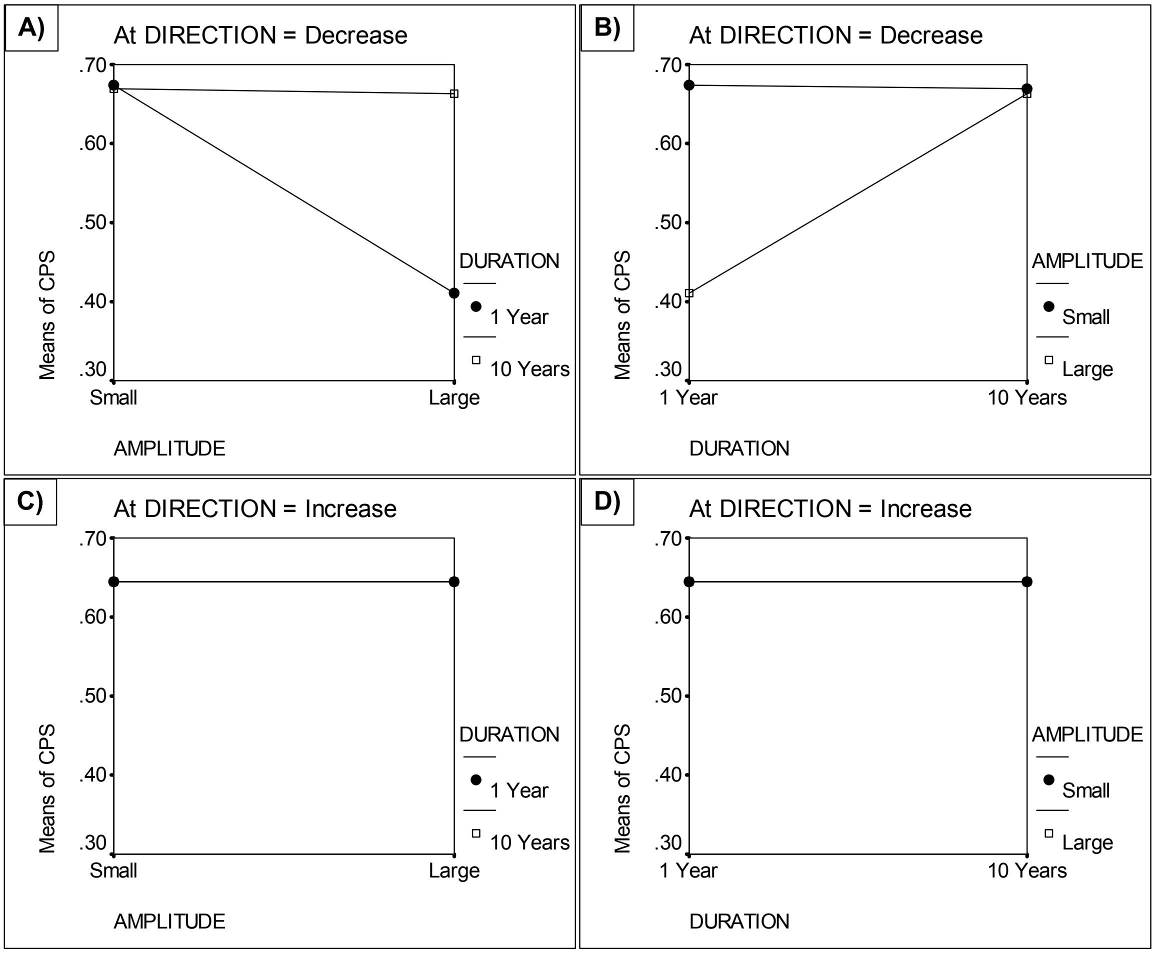


Figure 4.7: Graphs represent three-way interaction effect on CPS for Design A (future health state level = 0.70) for each level of direction of change from current health state to future health state

By examining the three-way interaction by level of direction of change from current health state to future health state when future health state is at level 0.70 (as in Figure 4.7), it can obviously be seen that duration and amplitude interact only when direction of change from current health state to future health state is *decrease* (Figure 4.7A and Figure 4.7B as opposed to Figure 4.7C and Figure 4.7D). Consider Figure 4.7A, when future health state is decreased from current health state with a small amplitude change, having current health state duration of 1-year or 10-year do not significantly affect CPS. They lead to approximately equal CPS (0.674 and 0.670, respectively). On the other hand, when future health state is decreased from current health state with a large amplitude change, current health state duration does matter. Current health state that has 1-year duration leads to significantly smaller CPS than current health state that has 10-year duration by an average of 0.253 differences.

By interpreting from another perspective (Figure 4.7B), when future health state is decreased from current health state, a change in CPS when amplitude of change between current health state and future health state is varied from small to large when current health state has 1-year duration is significantly different from the change that occurs when current health state has 10-year duration. When future health state is decreased from current health state that has 1-year duration, varying amplitude of change from small to large significantly decreases CPS by an average of 0.264. However, when future health state is decreased from current health state that has 10-year duration, CPS for future health state that has small amplitude of change from current health state and CPS for future health state that has large amplitude of change from current health state are approximately equal (0.670 and 0.663, respectively).

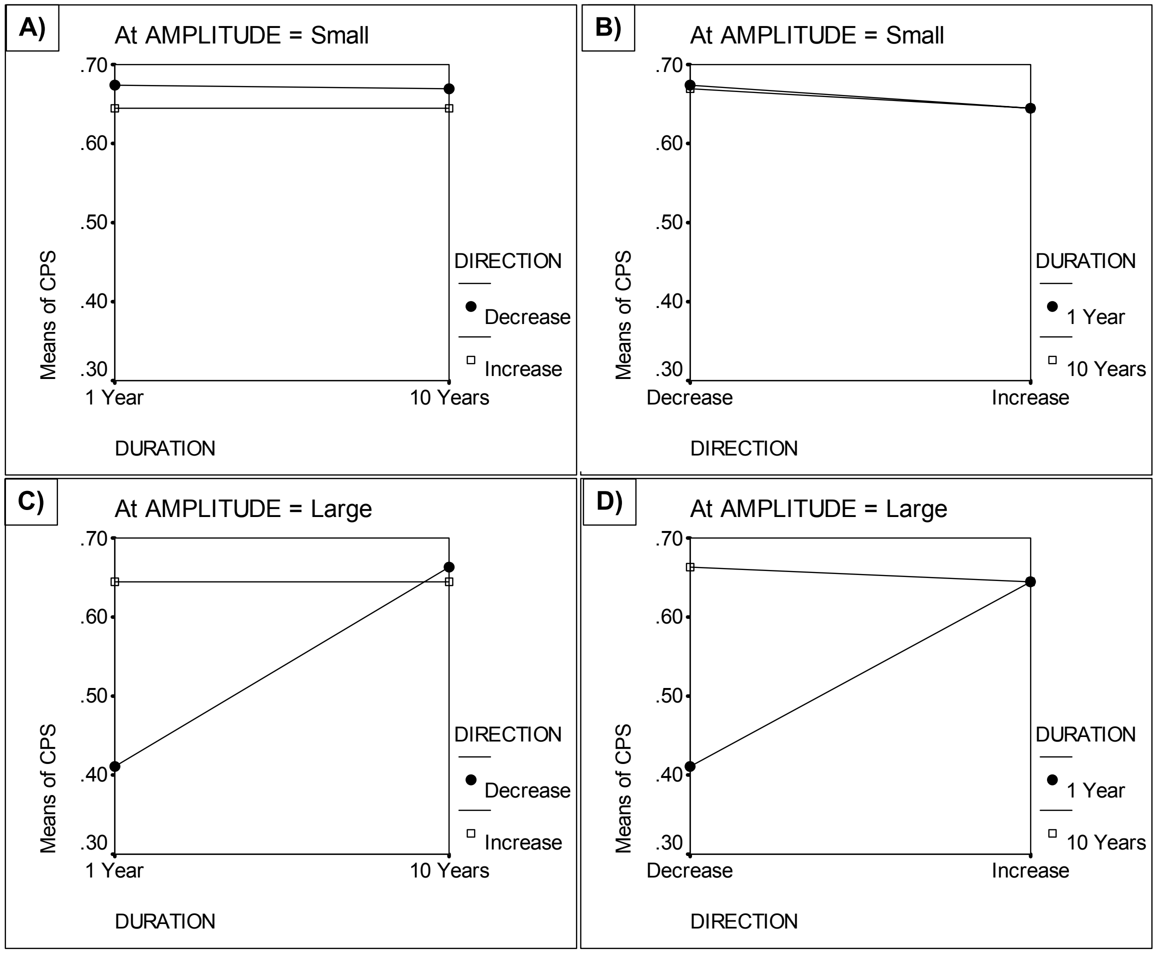


Figure 4.8: Graphs represent three-way interaction effect on CPS for Design A (future health state level = 0.70) for each level of amplitude of change between current health state and future health state

By examining the three-way interaction by level of amplitude of change between current health state and future health state when future health state is at level 0.70 (as in Figure 4.8), it can obviously be seen that direction and duration interact only when amplitude of change between current health state and future health state is large (Figure 4.8C and Figure 4.8D as opposed to Figure 4.8A and Figure 4.8B). Consider Figure 4.8C, when amplitude of change between current health state and future health state is large, a change in CPS when direction of change is varied from decreasing to increasing when current health state has 1-year duration is significantly different from the change when current health state has 10-year duration. When current health state has 1-year duration and amplitude of change between current health state and future health state is large, CPS for a future health state that is decreased from current health state is significantly lower than CPS for an increased future health state, with an average of 0.235 score lower. However, when current health state has 10-year duration and amplitude of change between current health state and future health state is large, CPS for a decreased future health state is slightly higher, by an average of 0.019, than CPS for an increased future health state.

From another perspective (Figure 4.8D), when amplitude of change between current health state and future health state is large, a change in CPS, when current health state duration is varied from 1 year to 10 years for decreased future health state, is significantly different from the change that occurs when future health state is increased from current health state. When a future health state is decreased with large amplitude of change, CPS for future health state when current health state has 1-year duration is significantly smaller than CPS for future health state when current health state has 10-year duration by an average of 0.253. However, when future health state is increased from current health state with large amplitude, CPS for future health state when current health state has 1-year duration and when current health state has 10-year duration are approximately equal (0.644 and 0.645, respectively).

Hypothesis 2: **There is a significant effect of interaction between direction and amplitude.**

Table 4.15 shows the summary of mean CPS for each combination of levels of direction and amplitude. Additionally, Table 4.16 summarizes the results from GLM-RM for testing the interaction effect between direction and amplitude. The results show that 2-way interaction effect between direction and amplitude is significant in all three designs with p-values less than 0.05. Figure 4.9, Figure 4.10, and Figure 4.11 depict interaction plots between direction and amplitude for Design A, B, and C respectively.

Table 4.15: Mean of CPS for each combination of direction and amplitude for each design

Direction	**Amplitude**	**Mean (Design A)**	**Mean (Design B)**	**Mean (Design C)**
Decrease	Small	0.672	0.535	0.413
	Large	0.537	0.542	0.413
Increase	Small	0.644	0.536	0.401
	Large	0.645	0.499	0.467

Table 4.16: Summary of the results from GLM-RM for 2-way interaction effect between direction and amplitude

2-Way Interaction Effect (direction x amplitude)	**GLM-RM**	
	Test statistic	p-value
Design A (Future health state level = 0.70)	F = 46.91	<0.001
Design B (Future health state level = 0.55)	F = 4.21	0.043
Design C (Future health state level = 0.40)	F = 9.47	0.003

*The grey highlight indicates significant effect at alpha = 0.05

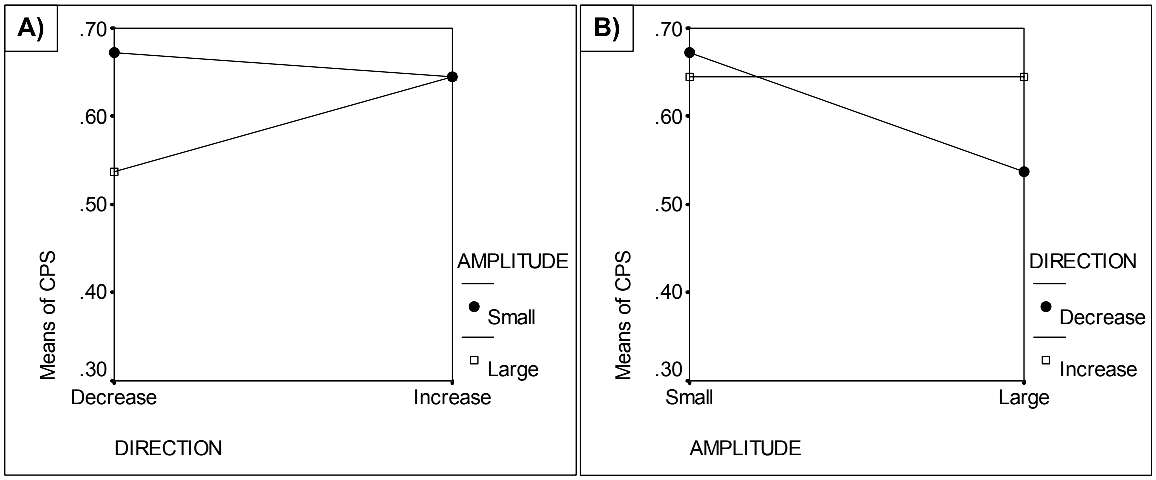


Figure 4.9: Interaction plots represent 2-way interaction effect between direction and amplitude on CPS for Design A (future health state level = 0.70)

From Figure 4.9A, when future health state is at level 0.70, the difference in CPS when the amplitude of change from current health state to future health state is varied from small to large for a decreased future health state is significantly different from that for an increased future health state. For future health state that is decreased from current health state, varying the amplitude of change from small to large significantly decreases CPS for future health state by an average of 0.135 score. However, when future health state is increased from current health state, CPS for future health state that has small amplitude of change from current health state and CPS for future health state that has large amplitude of change from current health state are approximately equal (0.644 and 0.645, respectively).

As shown in Figure 4.9B, change in CPS, when the direction of change from current health state to future health state is varied between decreasing and increasing and when the amplitude of change between current health state and future health state is small, is significantly different from the change when amplitude of change is large. When future health state has small amplitude of change from current health state, varying direction of change from decreasing to increasing slightly decreases CPS by an average of 0.028. On the other hand, when future health state has large amplitude of change from current health state, CPS for a decreased future health state is significantly lower than CPS for an increased future health state, with an average of 0.107 score of difference.

Based on the result of the significant 3-way interaction between duration, direction, and amplitude, the 2-way interaction between direction and amplitude, described above in terms of tendency of changes in CPS when different levels of direction and amplitude are varied, is present only when current health state has 1-year duration.

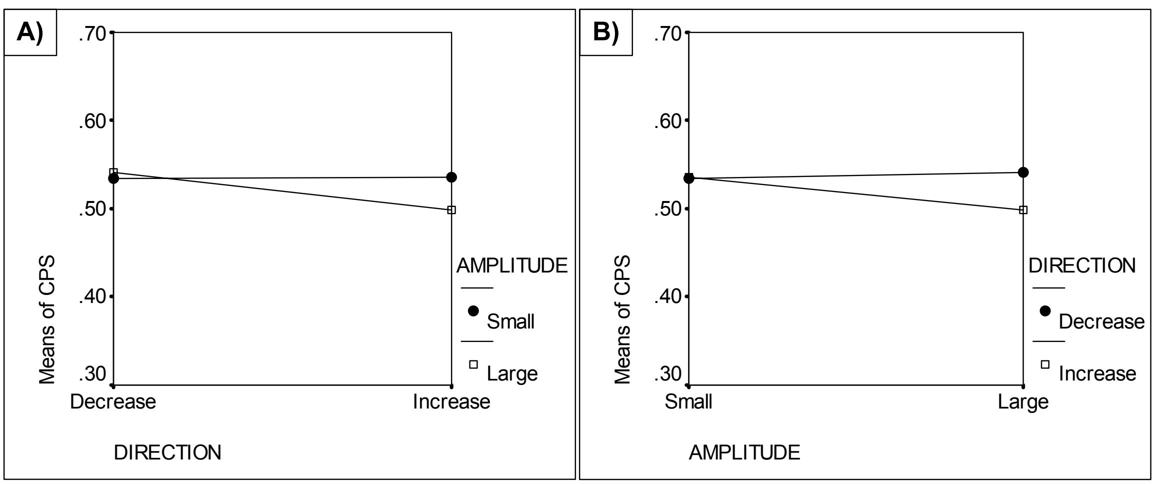


Figure 4.10: Interaction plots represent 2-way interaction effect between direction and amplitude on CPS for Design B (future health state level = 0.55)

As shown in Figure 4.10A, when future health state is at level 0.55, the difference in CPS when the amplitude of change from current health state to future health state is varied from small to large for a decreased future health state is significantly different from that of an increased future health state. When future health state is decreased from current health state, the CPS for the future health state that has large amplitude of change from current health state is slightly higher, with an average of 0.007, than the CPS for the future health state that has small amplitude of change. On the other hand, when future health state is increased from current health state, CPS for future health state that has large amplitude is lower than the CPS for future health state that has small amplitude of change from current health state by an average of 0.037 score.

Looking at Figure 4.10B, change in CPS, when the direction of change from current health state to future health state is varied between decreasing and increasing and when the amplitude of change between current health state and future health state is small, is significantly different from the change when amplitude of change is large. When future health state has small amplitude of change from current health state, CPS for a decreased future health state is approximately equal to CPS from an increased future health state (0.535 and 0.536, respectively). However, when future health state corresponds to a large amplitude change from current health state, CPS for a decreased future health state is higher than CPS for an increased future health state, with an average of 0.043 score of difference.

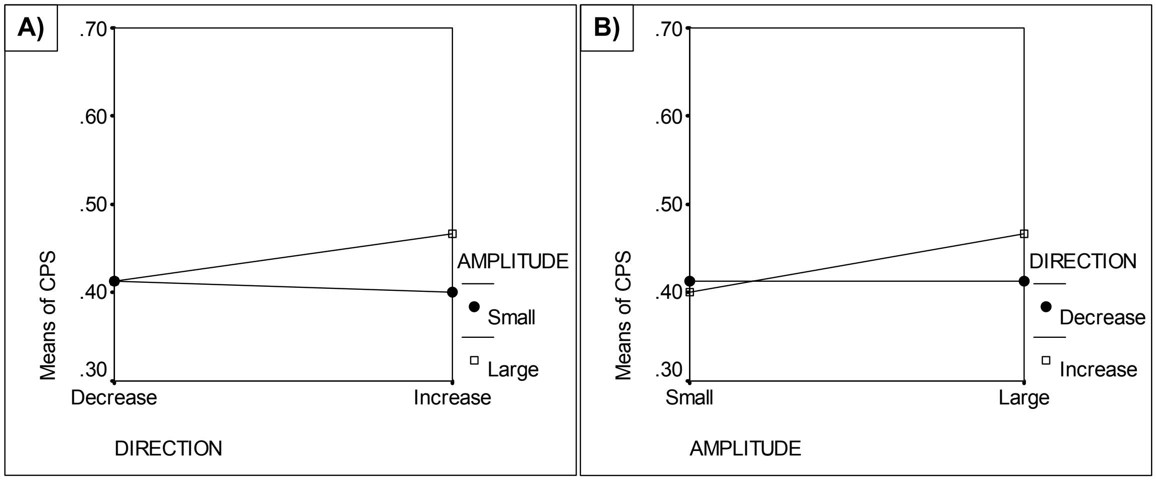


Figure 4.11: Interaction plots represent 2-way interaction effect between direction and amplitude on CPS for Design C (future health state level = 0.40)

From Figure 4.11A, when future health state is at level 0.40, the difference in CPS, when the amplitude of change from current health state to future health state is varied from small to large for a decreased future health state, is significantly different from that for an increased future health state. For a future health state that is decreased from current health state, CPS for future health state that corresponds to a large amplitude change from current health state is approximately equal to CPS for future health state that has small amplitude of change from current health state. However, for a future health state that is increased from current health state, varying the amplitude of change from small to large significantly increases CPS for future health state by an average of 0.066 score.

By looking at the issue from another perspective (Figure 4.11B), change in CPS, when the direction of change from current health state to future health state is varied between decreasing and increasing and when the amplitude of change between current health state and future health state is small, is significantly different from change when amplitude of change is large. When future health state corresponds to a small amplitude change from current health state, varying direction of change from decreasing to increasing slightly decreases CPS by an average of 0.012 score. On the other hand, when future health state corresponds to a large amplitude change from current health state, CPS for a decreased future health state is lower than CPS for an increased future health state, with an average of 0.054 score of difference.

Hypothesis 3: **There is no significant interaction effect between direction and duration on the conditional preference score of the future health state.**

Table 4.17 shows the summary of the mean CPS for each combination of levels between direction and duration. Furthermore, Table 4.18 summarizes the results from GLM-RM for testing the interaction effect between direction and duration. The results show that 2-way interaction effect between direction and amplitude is significant only in Design A with p-value of less than 0.001. Figure 4.12 depicts interaction plots between direction and duration for Design A.

Table 4.17: Mean of CPS for each combination of direction and duration for each design

Duration	Direction	Mean (Design A)	Mean (Design B)	Mean (Design C)
1 Year	Decrease	0.542	0.540	0.408
	Increase	0.645	0.507	0.418
10 Years	Decrease	0.666	0.537	0.418
	Increase	0.644	0.528	0.450

Table 4.18: Summary of the results from GLM-RM and FDA for 2-way interaction effect between direction and duration

2-Way Interaction Effect (direction x duration)	**GLM-RM**	
	Test statistic	p-value
Design A (Future health state level = 0.70)	F = 62.04	<0.001
Design B (Future health state level = 0.55)	F = 3.04	0.085
Design C (Future health state level = 0.40)	F = 1.58	0.212

*The grey highlight indicates significant effect at alpha = 0.05

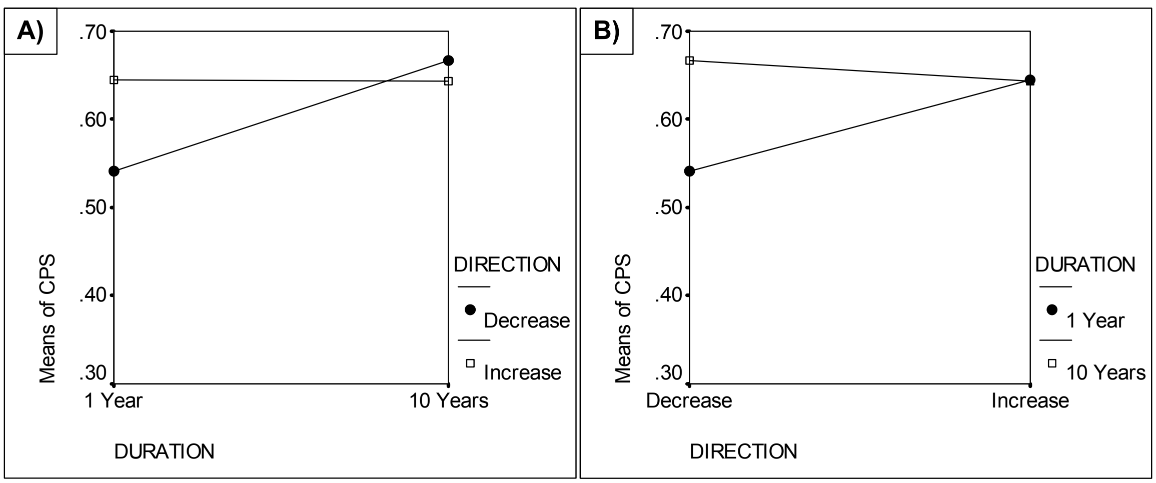


Figure 4.12: Interaction plots represent 2-way interaction effect between direction and duration on CPS for Design A (future health state level = 0.70)

From Figure 4.12A, it can be seen that, when future health state is at level 0.70, a change in CPS when direction of change from current health state to future health state is varied from decreasing to increasing and when current health state has 1-year duration, is significantly different from the corresponding change when current health state has 10-year duration. When current health state has 1-year duration, CPS for a future health state that is decreased from current health state is significantly lower than CPS for an increased future health state, by an average of 0.103. On the other hand, when current health state has 10-year duration, CPS for a decreased future health state tends to be higher by an average of 0.022 score, not significantly, than CPS for an increased future health state.

As seen in Figure 4.12B, a change in CPS, when current health state duration is varied from 1 year to 10 years for decreased future health state, is significantly different from the change when future health state is increased from current health state. When a future health state is decreased from the current health state, CPS for future health state when current health state has a 1-year duration is significantly smaller than CPS for future health state when current health state has a 10-year duration by an average of 0.124 score. However, when future health state is increased from current health state, means of CPS for future health state when current health state has a 1-year duration and when current health state has a 10-year duration are approximately equal (0.645 and 0.644, respectively).

Based on the results of the significant 3-way interaction between duration, direction, and amplitude, the 2-way interaction between duration and direction described previously (in terms of tendency of changes in CPS when different levels of duration and direction are varied) is present only when the amplitude of change from current health state to future health state is large.

Hypothesis 4: **There is no significant interaction effect between amplitude and duration on the conditional preference score of the future health state.**

Table 4.19 shows the summary of the mean CPS for each combination of levels between amplitude and duration. Furthermore, Table 4.20 summarizes the results from GLM-RM for testing the interaction effect between amplitude and duration. The results show that the 2-way interaction effect between amplitude and duration is significant only in Design A with p-value less than 0.001. Figure 4.13 depicts interaction plots between amplitude and duration for Design A.

Table 4.19: Mean of CPS for each combination of amplitude and duration for each design

Duration	**Amplitude**	**Mean (Design A)**	**Mean (Design B)**	**Mean (Design C)**
1 Year	Small	0.659	0.535	0.403
	Large	0.528	0.512	0.423
10 Years	Small	0.657	0.536	0.411
	Large	0.653	0.528	0.457

Table 4.20: Summary of the results from GLM-RM for 2-way interaction effect between amplitude and duration

2-Way Interaction Effect (amplitude x duration)	**GLM-RM**	
	Test statistic	p-value
Design A (Future health state level = 0.70)	F = 75.38	<0.001
Design B (Future health state level = 0.55)	F = 1.23	0.271
Design C (Future health state level = 0.40)	F = 2.72	0.102

*The grey highlight indicates significant effect at alpha = 0.05

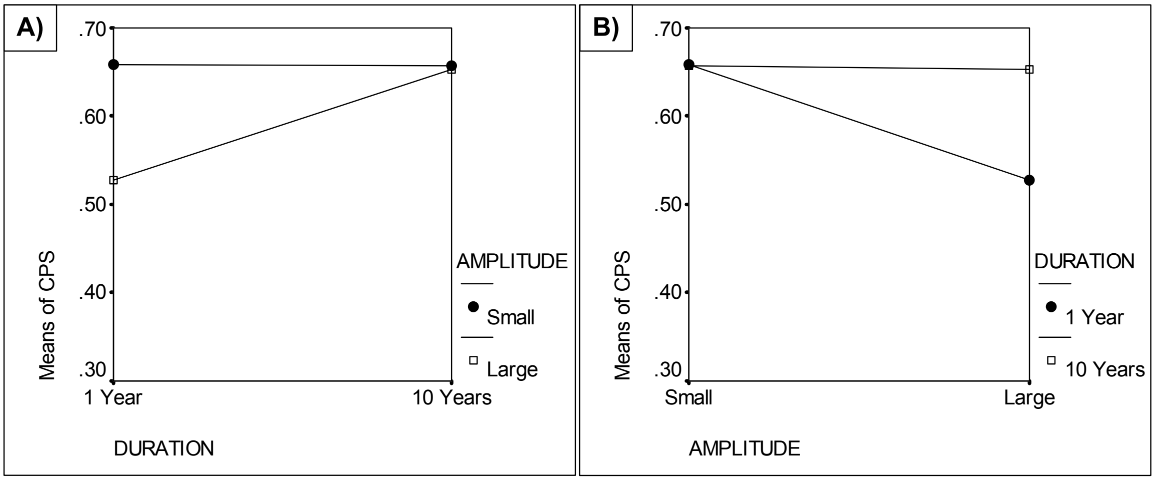


Figure 4.13: Interaction plots represent 2-way interaction effect between amplitude and duration on CPS for Design A (future health state level = 0.70)

As seen in Figure 4.13A, when future health state is at level 0.70, a change in CPS when the amplitude of change from current health state to future health state is varied from small to large and when current health state has a 1-year duration, is significantly different from the change when current health state has a 10-year duration. When current health state has a 1-year duration, varying amplitude of change from small to large significantly decreases CPS by an average score of 0.131. However, when current health state has a 10-year duration, mean of CPS for a future health state that has large amplitude of change from current health state and mean of CPS for a future health state that has small amplitude of change from current health state are approximately equal (0.657 and 0.653, respectively).

As shown in Figure 4.13B, a change in CPS when current health state duration is varied from 1 year to 10 years for future health state that has small amplitude of change is significantly different from the change when future health state has large amplitude of change. When future health state corresponds to a small amplitude change from current health state, the mean CPS for future health state when current health state has a 1-year duration is approximately equal to the mean CPS for future health state when current health state has a 10-year duration (0.659 and 0.657, respectively). Nonetheless, when future health state corresponds to a large amplitude change from current health state, CPS for future health state when current health state has a 1-year duration is significantly lower, by an average of 0.125, than CPS for future health state when current health state has 10-year duration.

Based on the result of the significant 3-way interaction between duration, direction, and amplitude, the 2-way interaction between duration and amplitude described above (in terms of tendency of changes in CPS when different levels of duration and amplitude are varied) is present only when future health state is decreased from current health state.

Hypothesis 5: **A main effect of duration will result in the interdependence between the future health state and the current health state. The effect will be less significant on preference when the current health state has shorter duration.**

This hypothesis states that current health state will have a stronger effect on future health state when current health state has longer duration than when it has shorter duration. One way to investigate this hypothesis is to look at the effects of other factors of current health state (direction and amplitude) on different levels of duration and analyze how each factor affects the conditional preference score. In this analysis, GLM-RM was performed for each design for each duration. Direction and amplitude are independent variables with conditional preference score as a dependent variable. Table 4.21 shows the results from GLM-RM.

Table 4.21: Results from GLM-RM by duration for each Design

Design	Factors	1-Year Duration		10-Year Duration	
		Statistic	p-value	Statistic	p-value
Design A (Future health state level = 0.70)	Direction	F = 37.41	<0.001	F = 2.09	0.152
	Amplitude	F = 111.18	<0.001	F = 0.11	0.737
	DirectionxAmplitude	F = 101.23	<0.001	F = 0.08	0.775
Design B (Future health state level = 0.55)	Direction	F = 5.71	0.019	F = 0.37	0.547
	Amplitude	F = 4.07	0.047	F = 046	0.499
	DirectionxAmplitude	F = 6.69	0.011	F = 0.89	0.347
Design C (Future health state level = 0.40)	Direction	F = 0.35	0.554	F = 4.83	0.030
	Amplitude	F = 2.98	0.088	F = 16.00	<0.001
	DirectionxAmplitude	F = 5.25	0.024	F = 7.61	0.007

*The grey highlight indicates significant effect at alpha = 0.05

Table 4.21 shows that direction, amplitude, and their interaction have all significant effects for scenarios with a 1-year current health state duration but not for those with a 10-year current health state duration for both Design A and B (p-values from less than 0.001 to 0.047). On the other hand, for Design C, all effects are significant for scenarios with the 10-year current health state duration (p-values from less than 0.001 to 0.030). However, for Design C, for scenarios with 1-year current health state duration, only the interaction effect is significant (p-value = 0.024).

The results thus lead to rejecting hypothesis#5 for Design A and B since they shows that effects of two different characteristics of the current health state (direction and amplitude) are more significant when the current health state duration is shorter.

However, the results for Design C are consistent with hypothesis#5 in that the effects are more significant when the current health state duration is longer.

An alternative approach for exploring and testing this hypothesis is to look at the deviation of the conditional preference score from the actual future health state level (unconditional preference score) for each current health state duration. For each subject, absolute values of the difference between the unconditional preference score of the future health state (e.g. at level 0.70, 0.55, or 0.40) and the conditional preference score for that future health state were calculated. Those values for scenarios that have a 1-year current health state duration were then compared to those that have 10-year current health state duration. Since the normality tests showed that the data is not normality distributed, non-parametric test (Wilcoxon signed ranks) was used for the statistical comparison. Descriptive statistics and test results are shown in Table 4.22.

Table 4.22: Descriptive statistics of deviation between the unconditional preference scores and the unconditional preference scores for each future health state level in each design and the results from Wilcoxon-Signed Ranks tests

Design	**Current Health State Duration**	**Min**	**Max**	**Mean**	**S.D.**	**Median**	**Wilcoxon Signed Ranks Tests**
A	1 year	0.000	0.679	0.150	0.133	0.110	Z = -3.173 p-value = 0.002
	10 years	0.000	0.475	0.121	0.093	0.097	
B	1 year	0.000	0.424	0.105	0.087	0.084	Z = -0.438 p-value = 0.661
	10 years	0.000	0.418	0.108	0.088	0.094	
C	1 year	0.001	0.593	0.127	0.107	0.102	Z = -1.298 p-value = 0.194
	10 years	0.001	0.652	0.136	0.121	0.109	

*The grey highlight indicates significant effect at alpha = 0.05

Table 4.22 shows that only Design A leads to a significant result in comparing the deviation between when the scenarios has 1-year current health state duration and when they have 10-year current health state duration. It shows that when current health state has a 1-year duration, the deviation from the actual future health state unconditional score is significantly higher than when current health state has a 10-year duration, indicating that current health state has a significantly larger effect on conditional preference score for future health state when the subjects have been experiencing current health state in shorter duration. This finding leads to rejecting hypothesis #5 for Design A. For Design B and C, the results are not significant.

Analysis of Main Effects

Besides testing the hypotheses, main effects of the factors were also further analyzed. Table 4.23 shows the summary of mean CPS for each level of duration. Table 4.24 summarizes the results from GLM-RM for testing the main effect of duration. The results show that the main effect of duration is significant in Design A (p-value < 0.001) and Design C (p-value = 0.013). Figure 4.14 and Figure 4.15 depict plots of duration for Design A and C, respectively.

Table 4.23: Mean of CPS for each level of duration for each design

Duration	**Mean (Design A)**	**Mean (Design B)**	**Mean (Design C)**
1 Year	0.593	0.523	0.413
10 Years	0.655	0.532	0.434

Table 4.24: Summary of the results from GLM-RM for main effect of duration

Duration Main Effect	GLM-RM	
	Test statistic	p-value
Design A (Future health state level = 0.70)	F = 47.24	<0.001
Design B (Future health state level = 0.55)	F = 1.34	0.250
Design C (Future health state level = 0.40)	F = 6.42	0.013

*The grey highlight indicates significant effect at alpha = 0.05

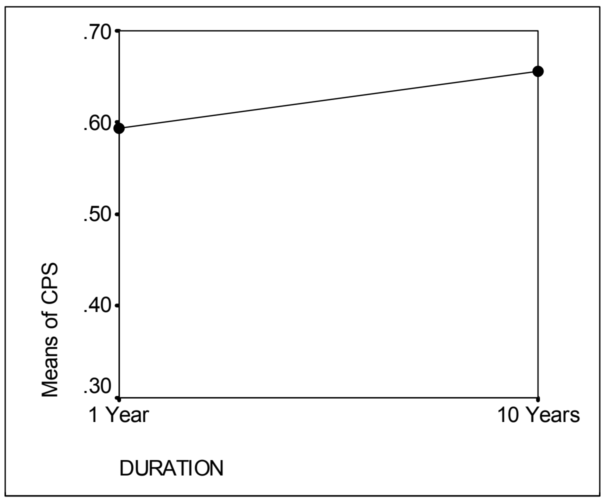


Figure 4.14: A plot represents main effect of duration on CPS for Design A (future health state level = 0.70)

From Figure 4.14, it can be seen that for future health state at level 0.70, CPS for future health state when current health state has a 1-year duration is significantly lower than CPS for future health state when current health state has a 10-year duration; with an average of 0.062 score of difference. However, duration effect also interacts with direction and amplitude. Thus, interpretation by considering an effect of duration solely is not appropriate.

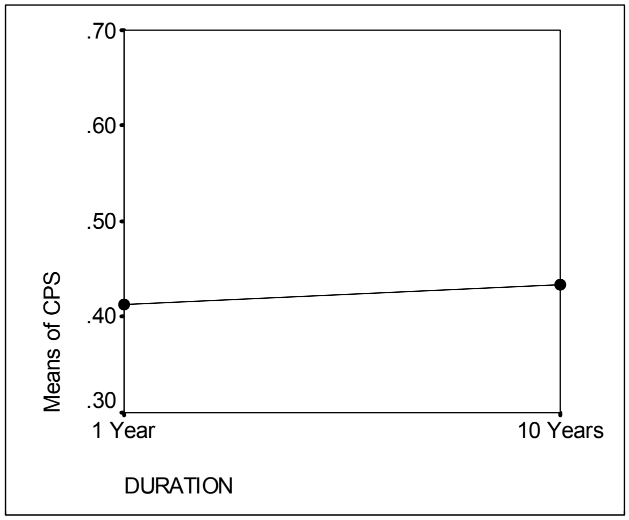


Figure 4.15: A plot represents main effect of duration on CPS for Design C (future health state level = 0.40)

From Figure 4.15, for future health state at level 0.40, CPS for future health state when current health state has a 1-year duration is significantly lower than CPS for future health state when current health state has a 10-year duration; with an average of 0.021 score of difference. However, duration effect also interacts with direction and amplitude. Thus, interpretation by considering an effect of duration solely is not appropriate.

Main effects of direction and amplitude were also analyzed, although it has been hypothesized that direction and amplitude interacts with each other and thus the main effects can be ignored. Table 4.25 shows the summary of mean CPS for each level of direction. Table 4.26 summarizes the results from GLM-RM for testing the main effect of direction. The results show that the main effect of direction is significant only in Design A with p-value of 0.005. Figure 4.16 illustrates a plot of direction for Design A.

Table 4.25: Mean of CPS for each level of direction for each design

Direction	Mean (Design A)	Mean (Design B)	Mean (Design C)
Decrease	0.604	0.538	0.413
Increase	0.644	0.517	0.434

Table 4.26: Summary of the results from GLM-RM for main effect of direction

Direction Main Effect	GLM-RM	
	Test statistic	p-value
Design A (Future health state level = 0.70)	F = 8.11	0.005
Design B (Future health state level = 0.55)	F = 2.87	0.094
Design C (Future health state level = 0.40)	F = 2.65	0.107

*The grey highlight indicates significant effect at alpha = 0.05

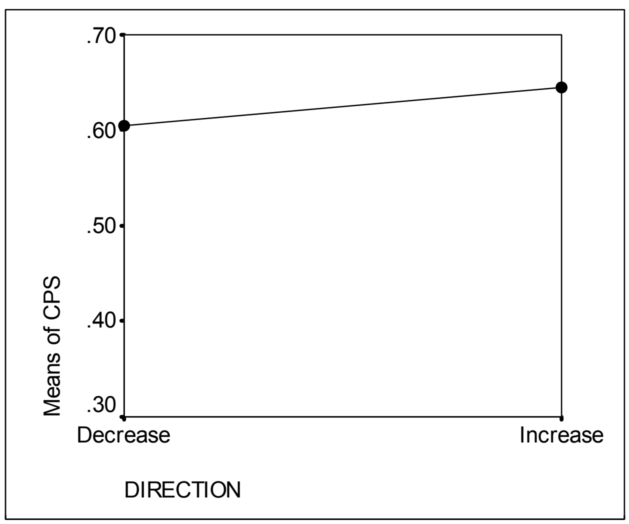


Figure 4.16: A plot represents main effect of direction on CPS for Design A (future health state level = 0.70)

From Figure 4.16, it can be seen that for future health state at level 0.70, CPS for a decreased future health state is significantly lower than CPS for an increased future health state; with an average score difference of 0.040. However, direction also interacts with duration and amplitude. Thus, interpretation by considering an effect of direction solely is not appropriate.

Regarding the amplitude main effect, Table 4.27 shows the summary of mean CPS for each level of amplitude. Table 4.28 summarizes the results from GLM-RM for testing a main effect of amplitude. The results show that the main effect of amplitude is significant in Design A and C with p-values less than 0.001. Figure 4.17 and Figure 4.18 depict plots of amplitude for Design A and C, respectively.

Table 4.27: Mean of CPS for each level of amplitude for each design

Amplitude	**Mean (Design A)**	**Mean (Design B)**	**Mean (Design C)**
Small	0.658	0.535	0.407
Large	0.591	0.520	0.440

Table 4.28: Summary of the results from GLM-RM and FDA for main effect of amplitude

Amplitude Main Effect	**GLM-RM**	
	Test statistic	p-value
Design A (Future health state level = 0.70)	F = 58.31	<0.001
Design B (Future health state level = 0.55)	F = 2.96	0.089
Design C (Future health state level = 0.40)	F = 14.48	<0.001

*The grey highlight indicates significant effect at alpha = 0.05

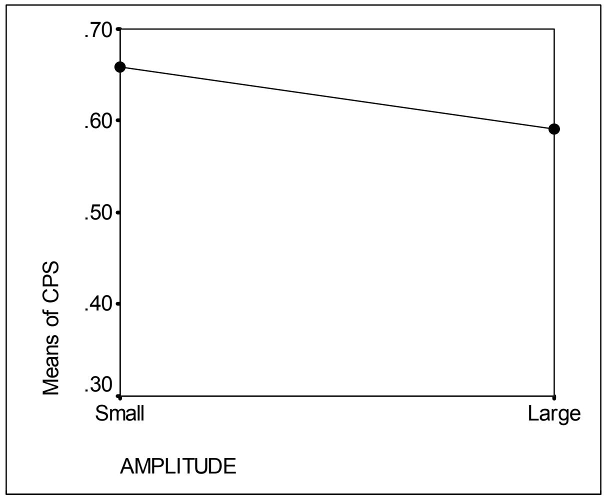


Figure 4.17: A plot represents main effect of amplitude on CPS for Design A (future health state level = 0.70)

From Figure 4.17, it can be seen that for future health state at level 0.70, CPS for a future health state that has small amplitude of change from current health state is significantly higher than CPS for a future health state that has large amplitude of change from current health state; with an average difference of 0.067. However, amplitude also interacts with duration and direction. Thus, interpretation by considering an effect of amplitude solely is not appropriate.

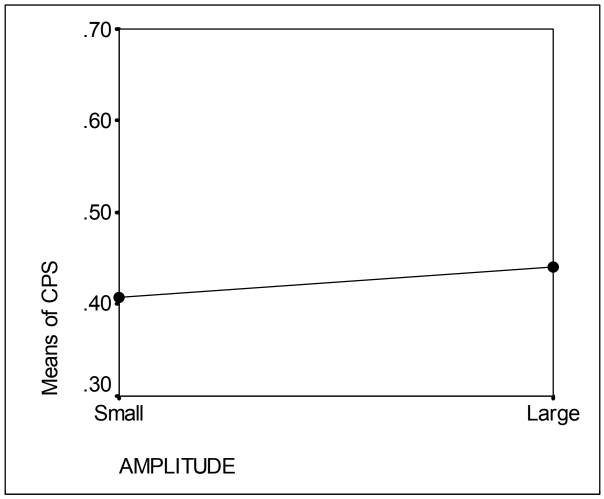


Figure 4.18: A plot represents main effect of amplitude on CPS for Design C (future health state level = 0.40)

From Figure 4.18, it can be seen that for future health state at level 0.40, CPS for a future health state that has small amplitude of change from current health state is significantly lower than CPS for a future health state that has large amplitude of change from current health state; with an average difference of 0.033. However, amplitude also interacts with duration and direction. Thus, interpretation by considering an effect of amplitude solely is not appropriate.

Summary of the Results

Table 4.29 summarizes the results of the analyses performed for each design to test each hypothesis. By looking at each hypothesis one by one, it appears that only hypothesis #2 (which indicates a significant 2-way interaction effect for direction and amplitude) has been confirmed for all three designs. Hypotheses #1, 3, and 4 are rejected by the results for Design A (future health state at 0.70 level). Regarding hypothesis #5, the first approach showed that only the results for Design C support the hypothesis. However, another approach showed that the results for all three designs lead to rejecting the hypothesis.

By looking at the results for each design, it is clearly seen that four of the five hypotheses are rejected by the results in Design A (future health state at 0.70 level); while for Design B (future health state at 0.55 level) and C (future health state at 0.40 level), the results generally confirm the hypotheses. Further discussion of those results will be provided in Chapter 5.

Table 4.29: Hypotheses and summary of the results in Phase 1

Hypotheses	Design A	Design B	Design C	Conclusions
1. There is NO significant three-way interaction effect between direction, amplitude and duration on the conditional preference score of the future health state.	×	✓	✓	The hypothesis is rejected only in Design A which is at level 0.70 of future health state.
2. There is a significant effect of interaction between direction and amplitude.	✓	✓	✓	The hypothesis is confirmed by the results from all three designs.
3. There is NO significant interaction effect between direction and duration on the conditional preference score of the future health state.	×	✓	✓	The hypothesis is rejected only in Design A which is at level 0.70 of future health state.
4. There is no significant interaction effect between amplitude and duration on the conditional preference score of the future health state.	×	✓	✓	The hypothesis is rejected only in Design A which is at level 0.70 of future health state.
*Hypothesis 5 used two approaches 5. A main effect of duration will result in the interdependence between the future health state and the current health state. The effect of duration will be less significant on preference when the current health state has shorter duration.	×	×	✓	The hypothesis is rejected by the results in Design A and B.
	×	×	×	The hypothesis is rejected by the results from all three designs.

4.3 Analysis and Results for Phase 2 Experiment

Phase 2 experiment aims at examining whether the conditional preference scores for discrete health state can better predict preferences scores for an entire multistate health profile than unconditional health state preference assessments. A total of ten health profiles were constructed, each composed of four consecutive health states. Eight of the 10 health profiles were assembled by varying three factors in a 2^3 factorial design. Those three factors are direction (2 levels: improvement or deterioration over time), duration (2 levels: 4year total duration versus 40-year or lifetime duration), and rate of change in health status over time (2 levels: gradual and steep). Another two health profiles were created and display no systematic pattern, one for each duration level. Table 4.30 summarizes key characteristics of each health profile. The graphical representations of the profiles are shown in Figures 3.14 to 3.23 in Chapter 3. As explained in Chapter 3, the three assessment techniques, conditional, unconditional, and holistic assessments are compared after duration-weighted scores are calculated from scores obtained from each assessment technique for each subject for each health profile.

Table 4.30: Summary of characterization of each health profile used in Phase 2 experiment

Health Profiles#	Duration	Slope	Direction
1	4 Years	Gradual	Decrease
2	4 Years	Gradual	Increase
3	4 Years	Steep	Decrease
4	4 Years	Steep	Increase
5	4 Years	No systematic pattern	
6	Lifetime (40 years)	Gradual	Decrease
7	Lifetime (40 years)	Gradual	Increase
8	Lifetime (40 years)	Steep	Decrease
9	Lifetime (40 years)	Steep	Increase
10	Lifetime (40 years)	No systematic pattern	

4.3.2 Data Preparation

Data that was used for Phase 2 analysis consisted of the 31 conditional preference scores, 6 unconditional preference scores, and 10 holistic preference scores. Conditional and unconditional preference scores were linearly normalized using mapping best health state (perfect health) and worse health state (death), as indicated previously in the stage of unconditional preference assessments, to 1.00 and 0.00, respectively. Eleven cases, which contained normalized scores that were higher than 1.00 or lower than 0.00, were removed. Subsequently, an outlier analysis was performed using box plot analysis. As a result, one response, identified as an extreme outlier, was removed. Box plots and descriptive statistics of the subjects whose data were removed for Phase 2 analysis are provided in Appendix C. Thus, data that was used in Phase 2 analysis correspond to 87 subjects.

4.3.3 Data Description

<u>Background Variables</u>

Phase 2 analysis used a total of 87 subjects, 32 (37%) female students and 55 (63%) male students. Age ranges from 18 to 33 years, with an average of 22.29 years. Among the 92 students, 43 (49%) are white, 24 (36%) are Asian or Pacific Islander, 10 (11%) are Hispanic, and 6 (7%) are black. The majority (79 (91%)) of them major in Industrial Engineering and 69 (79%) are undergraduate students. Concerning major health issues, 12 (14%) experienced one themselves, 46 (53%) experienced major health issues in their families, and 39 (45%) experienced health issues in significant others. Regarding the students health status, none of them indicates having poor health at the time of the experiment. A summary of the background variables of the 87 subjects is shown in Table 4.31. Figure 4.19 shows the distribution of age, the only continuous variable.

Table 4.31: Summary of Background Variables for Phase 2 Analysis

Variable	**Counts**	**Percent**	**Mean**	**S.D.**	**Range**
Gender					
Female	32	36.78			
Male	55	63.22			
Ethnic Background					
White, not of Hispanic origin	43	49.42			
Black, not of Hispanic origin	6	6.90			
Hispanic	10	11.49			
Asian or Pacific Islander	24	27.59			
Others	4	4.60			
Age (years)			22.29	2.46	18.00 - 33.00
Major field of study					
Industrial Engineering	79	90.80			
Other – non-engineering	6	6.90			
Other – Engineering	2	2.30			
Degree of study					
B.S.	69	79.31			
M.S.	7	8.05			
Ph.D.	10	11.49			
Other	1	1.15			
Health problem in yourself					
Yes, major issue(s)	12	13.79			
Yes, minor issue(s)	40	45.98			
No	35	40.23			
Health problem in your family					
Yes, major issue(s)	46	52.87			
Yes, minor issue(s)	28	32.18			
No	13	14.95			
Health problem in someone else close to you					
Yes, major issue(s)	39	44.83			
Yes, minor issue(s)	23	26.44			
No	36	28.73			
General health condition					
Excellent	25	28.74			
Very good	47	54.02			
Good	14	16.09			
Fair	1	1.15			
Poor	0	0.00			

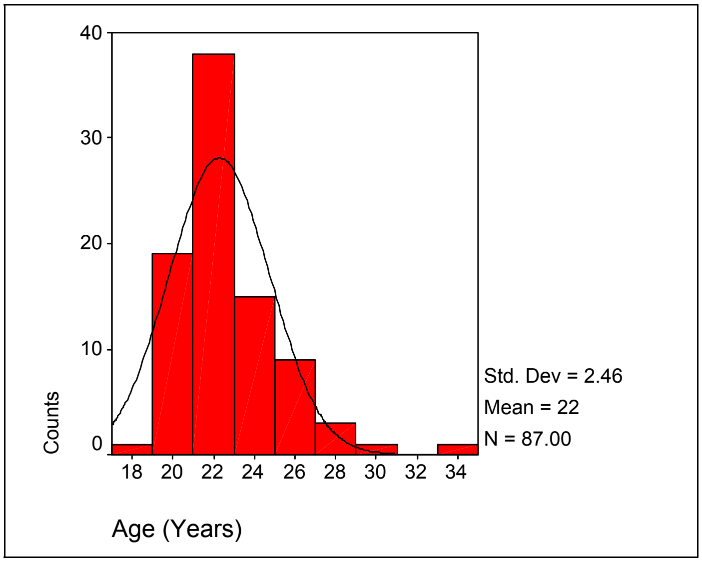


Figure 4.19: Age distribution of subjects in Phase 2 analysis

Holistic Preference Score

For each subject, a holistic preference score was obtained for each health profile by asking a subject to rate each health profile on a visual analog scale (VAS). Thus, each subject generated 10 holistic preference scores for a total of 10 health profiles. Table 4.32 shows descriptive statistics of holistic preference scores for each health profile obtained from 87 subjects. Notice that for the 4-year health profiles (profiles#1 to 5), increasing profiles have higher average scores than decreasing profiles. Health profile #2, which has a gradually increasing pattern, has an average score of 0.778 as opposed to 0.591 for health profile #1, which has a gradually decreasing pattern. Additionally, health profile #4, which has steeply increasing pattern has an average score of 0.572 as opposed to 0.206 for health profile #3, which has a steeply decreasing pattern. On the other hand, for the lifetime health profiles (profiles #6 to 10), decreasing profiles have higher average scores than increasing profiles. Health profile #6, which has a gradually decreasing pattern has an average score of 0.756 as opposed to 0.655 for health profile #7, which has a gradually increasing pattern. Additionally, health profile #8, which has a steeply decreasing pattern has an average score of 0.466 as opposed to 0.271 for health profile #8, which has a steeply increasing pattern.

Table 4.32: Holistic Preference Scores for Each Health Profile

Holistic Preference Scores	**N**	**Min**	**Max**	**Mean**	**S.D.**
Health Profile#1 (gradually decrease, 4 yr)	87	0.048	0.901	0.591	0.188
Health Profile#2 (gradually increase, 4 yr)	87	0.395	1.000	0.778	0.136
Health Profile#3 (steeply decrease, 4 yr)	87	0.000	0.937	0.206	0.174
Health Profile#4 (steeply increase, 4 yr)	87	0.053	1.000	0.572	0.249
Health Profile#5 (no pattern, 4 yr)	87	0.013	0.950	0.421	0.194
Health Profile#6 (gradually decrease, lifetime)	87	0.286	0.956	0.756	0.157
Health Profile#7 (gradually increase, lifetime)	87	0.227	0.958	0.655	0.172
Health Profile#8 (steeply decrease, lifetime)	87	0.017	0.956	0.466	0.238
Health Profile#9 (steeply increase, lifetime)	87	0.004	0.807	0.271	0.193
Health Profile#10 (no pattern, lifetime)	87	0.139	0.962	0.481	0.211

Duration-Weighted Scores (DWS)

As mentioned earlier in Chapter 3, final data that are used for the comparisons between the three assessment techniques (unconditional, conditional and holistic preference assessments) are duration-weighted. An example of duration-weighted scores (DWS) calculation for each assessment technique is shown in Chapter 3 (section 3.5.2).

Table 4.33 shows descriptive statistics of DWS obtained from 87 subjects from unconditional, conditional, and holistic preference assessment techniques for each health profile, along with the results from Kolmogorov-Smirnov normality tests. For 4-year health profiles (Profiles 1 to 5), means of DWS are ranging from 2.113 to 2.421 for unconditional preference assessment, from 1.910 to 2.618 for conditional preference assessment, and from 0.823 to 3.110 for holistic preference assessment. For lifetime health profiles (Profiles 6 to 10), means DWS are ranging from 17.866 to 24.214 for unconditional preference assessment, from 18.496 to 24.716 for conditional preference assessment, and from 10.884 to 30.240 for holistic preference assessment. The results from Kolmogorov-Smirnov normality tests show that for unconditional DWS, only the score for health profile #10 violates the normality assumption (p-value = 0.028). For conditional DWS, only the score for health profile #5 violates the normality assumption (p-value = 0.003). For holistic DWS, scores for health profiles #1, 2, 3, 6, and 9 do not exhibit normality (p-values from less than 0.001 to 0.003). Thus, for the analyses comparing DWS across techniques, non-parametric tests are used for profiles #1, 2, 3, 5, 6, 9, and 10, while parametric tests are used for profiles #4, 7, and 8. Figure 4.20 to Figure 4.29 illustrate the distributions of DWS from unconditional, conditional, and holistic preference assessment techniques for each health profile.

Table 4.33: DWS for each health profile for each assessment technique

Assessment Technique	**N**	**Min**	**Max**	**Mean**	**S.D.**	**Test of Normality (Kolmogorov-Smirnov)**	
						Statistics	**p-value**
Health Profile#1							
Conditional	87	1.799	3.010	2.465	0.227	0.063	0.200
Unconditional	87	2.237	2.635	2.421	0.085	0.086	0.159
Holistic	87	0.193	3.605	2.364	0.750	0.110	0.011
Health Profile#2							
Conditional	87	1.882	2.950	2.338	0.219	0.052	0.200
Unconditional	87	2.237	2.635	2.421	0.085	0.086	0.159
Holistic	87	1.580	4.000	3.110	0.543	0.160	0.000
Health Profile#3							
Conditional	87	1.275	2.737	1.910	0.301	0.090	0.081
Unconditional	87	1.938	2.333	2.113	0.086	0.076	0.200
Holistic	87	0.000	3.748	0.823	0.696	0.182	0.000

Table 4.33 (continued): DWS for each health profile for each assessment technique

Assessment Technique	N	Min	Max	Mean	S.D.	Test of Normality (Kolmogorov-Smirnov)	
						Statistics	p-value
Health Profile#4							
Conditional	87	1.470	2.884	2.143	0.264	0.081	0.200
Unconditional	87	1.938	2.333	2.113	0.086	0.076	0.200
Holistic	87	0.210	4.000	2.286	0.995	0.092	0.064
Health Profile#5							
Conditional	87	1.944	3.131	2.618	0.224	0.121	0.003
Unconditional	87	2.060	2.482	2.263	0.093	0.094	0.055
Holistic	87	0.050	3.798	1.683	0.775	0.061	0.200
Health Profile#6							
Conditional	87	18.631	30.770	24.716	2.120	0.064	0.200
Unconditional	87	22.371	26.354	24.214	0.855	0.086	0.159
Holistic	87	11.428	38.236	30.240	6.273	0.150	0.000
Health Profile#7							
Conditional	87	17.428	30.630	23.537	2.438	0.063	0.200
Unconditional	87	22.371	26.354	24.214	0.855	0.086	0.159
Holistic	87	9.076	38.320	26.214	6.875	0.071	0.200
Health Profile#8							
Conditional	87	13.487	27.135	21.738	2.404	0.061	0.200
Unconditional	87	19.381	23.333	21.127	0.859	0.076	0.200
Holistic	87	0.672	38.236	18.643	9.509	0.073	0.200
Health Profile#9							
Conditional	87	15.578	29.287	22.051	2.576	0.056	0.200
Unconditional	87	19.381	23.333	21.127	0.859	0.076	0.200
Holistic	87	0.168	32.268	10.844	7.708	0.122	0.003
Health Profile#10							
Conditional	87	12.416	28.109	18.496	2.476	0.065	0.200
Unconditional	87	15.773	20.444	17.866	1.087	0.101	0.028
Holistic	87	5.548	38.488	19.226	8.446	0.076	0.200

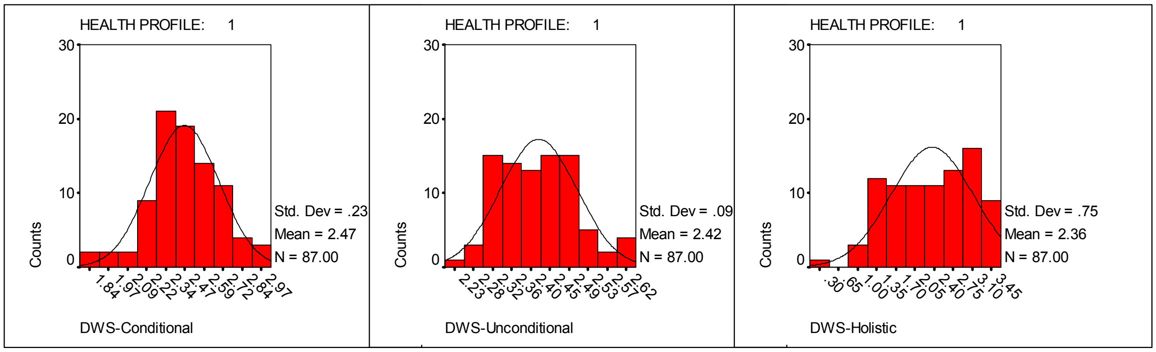


Figure 4.20: Distribution of DWS for health profile#1

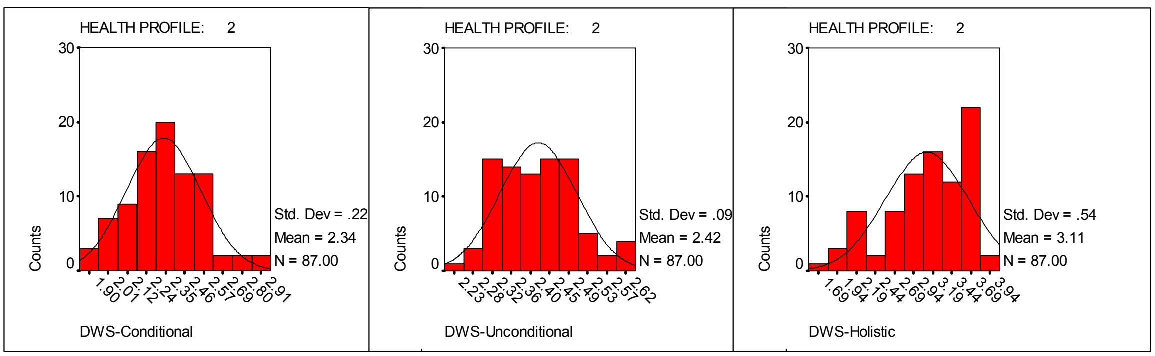


Figure 4.21: Distribution of DWS for health profile#2

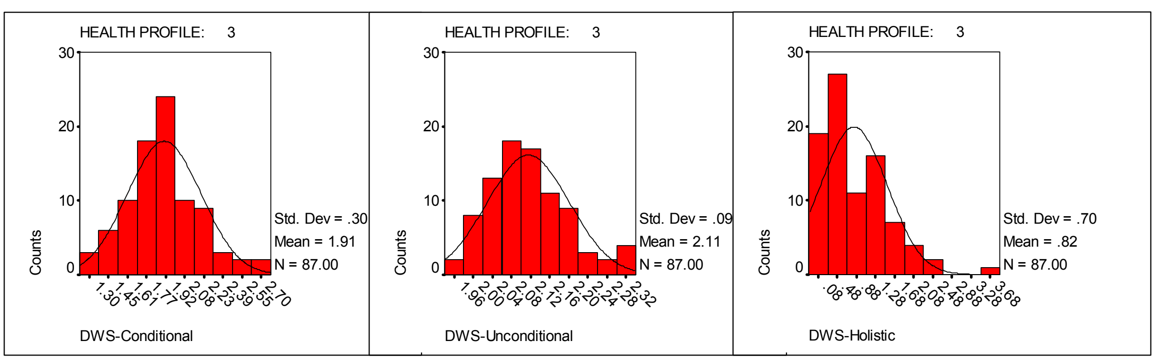


Figure 4.22: Distribution of DWS for health profile#3

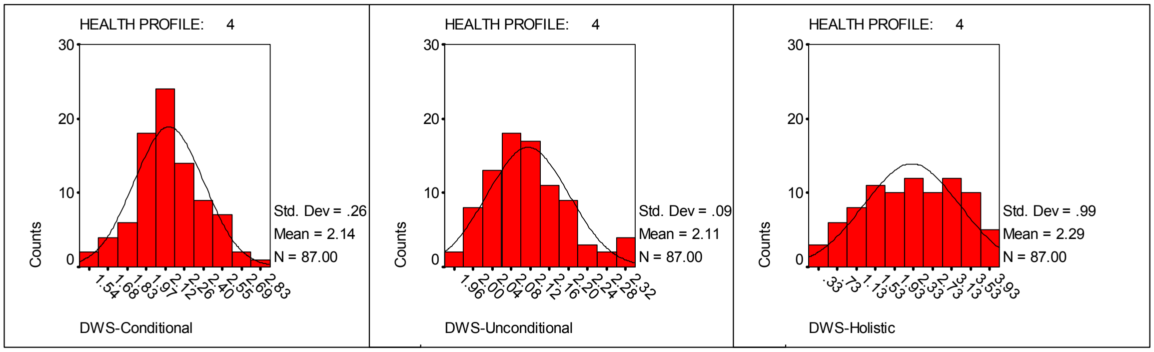


Figure 4.23: Distribution of DWS for health profile#4

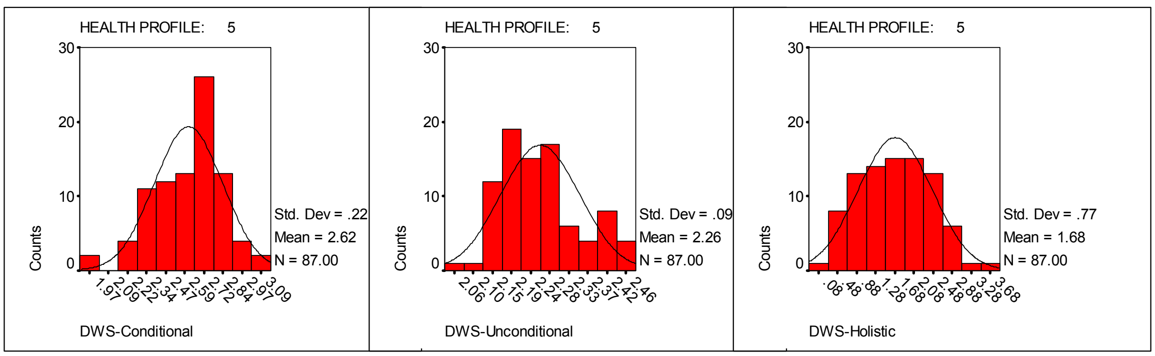


Figure 4.24: Distribution of DWS for health profile#5

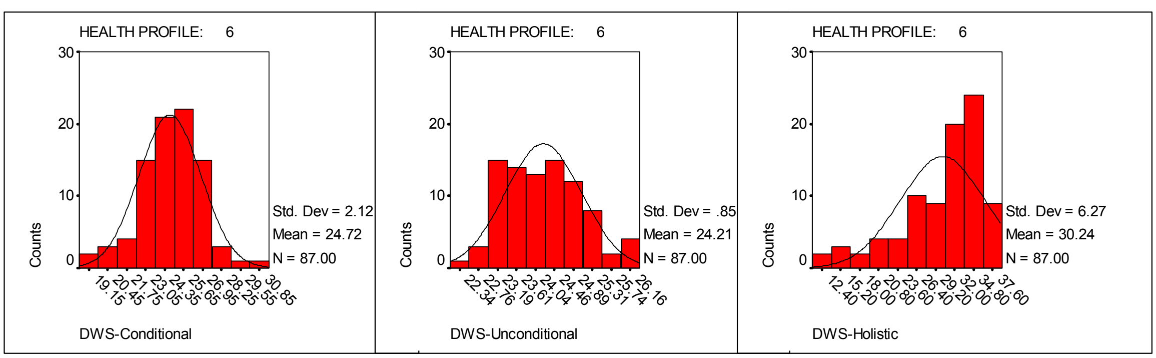


Figure 4.25: Distribution of DWS for health profile#6

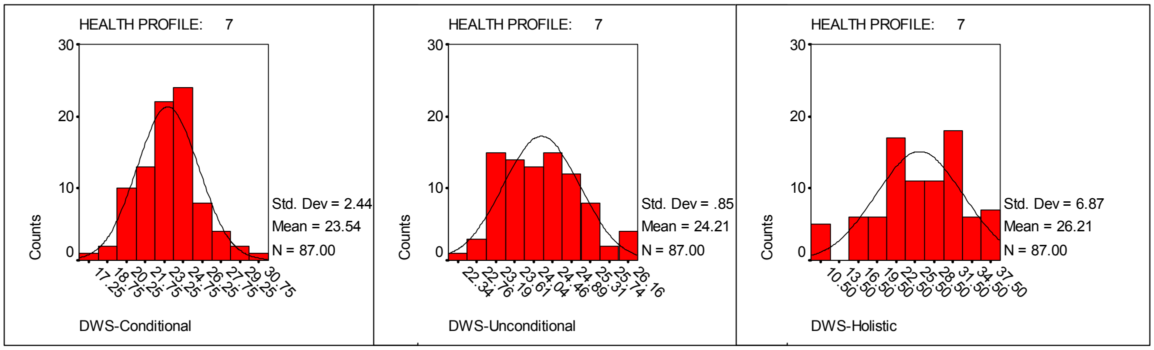


Figure 4.26: Distribution of DWS for health profile#7

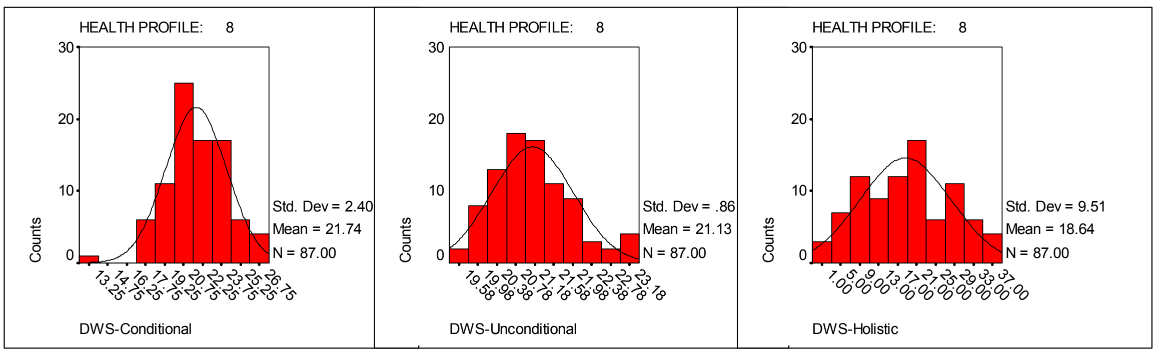


Figure 4.27: Distribution of DWS for health profile#8

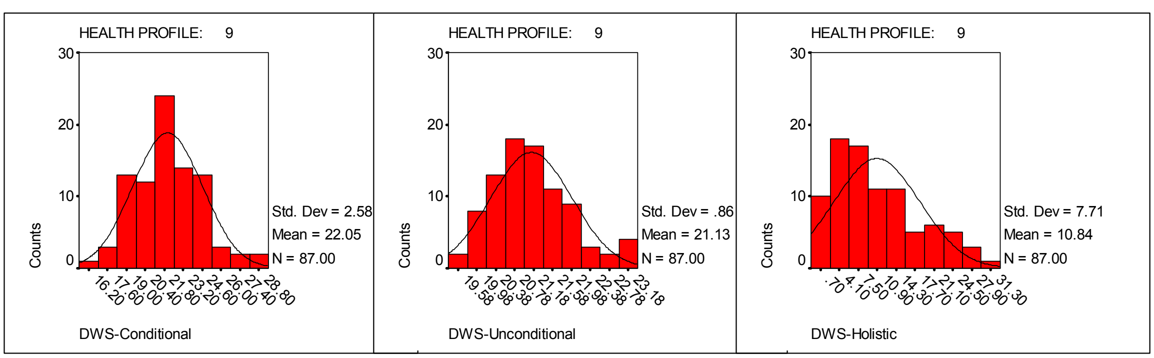


Figure 4.28: Distribution of DWS for health profile#9

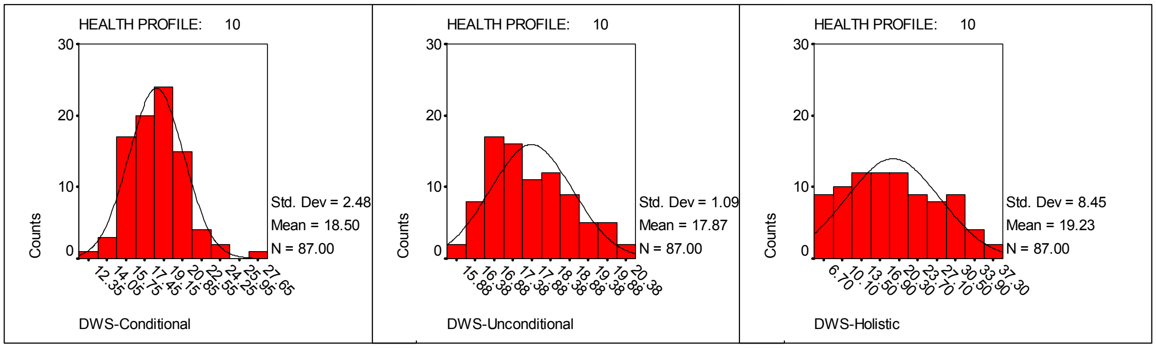


Figure 4.29: Distribution of DWS for health profile#10

4.3.4 Data Analysis

Data analysis mainly involved comparisons of duration-weighted scores across conditional, unconditional, and holistic preference assessment. However, normality tests revealed that normality assumption was not met in some health profiles. Thus, both non-parametric and parametric analyses had been used in this stage. Nonparametric analysis used Friedman tests with Wilcoxon Signed Rank tests for post-hoc tests, whereas parametric analysis employed GLM-RM with Bonferroni adjustment for post-hoc tests.

Hypothesis 6: **The proposed decomposition method in which several conditional preference score assessments are executed and integrated to produce a preference score for the entire health profile will predict the preference of the entire health profile better than the decomposition technique that uses unconditional preference score assessments.**

Figure 4.30 and Figure 4.31 depict mean duration-weighted scores across subjects for each preference assessment technique for 4-year health profiles (profiles #1 to 5) and for lifetime health profiles (profiles #6 to 10), respectively. Both figures show that DWS-conditional and DWS-unconditional are similar to each other; however, they both are quite different from DWS-holistic especially in some health profiles (profiles #2, 3, 6 and 9).

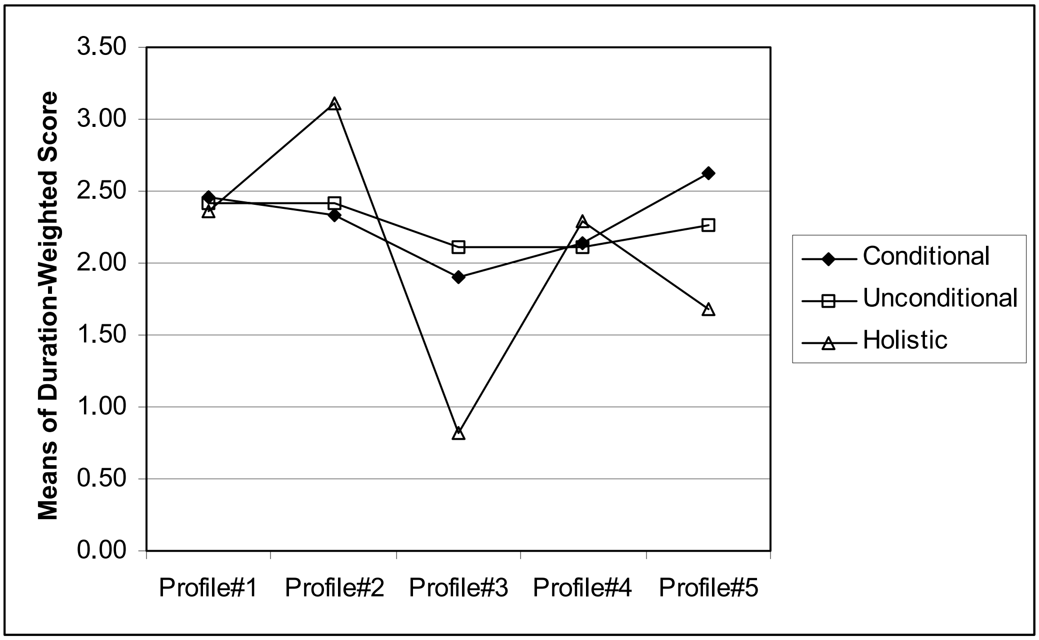


Figure 4.30: Graphs illustrate means of DWS from each assessment technique for health profiles 1 to 5

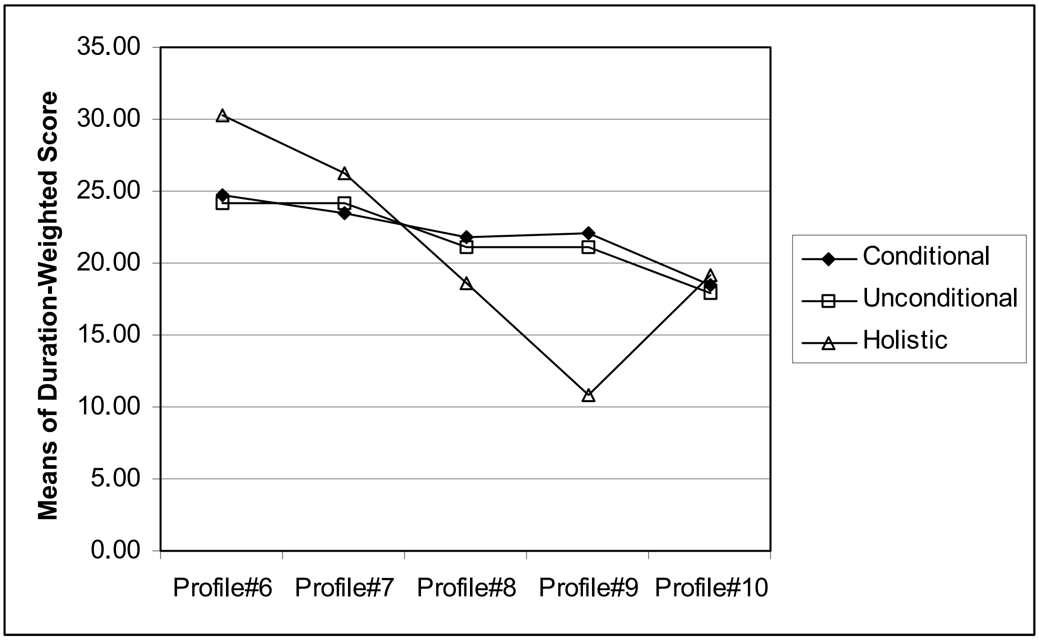


Figure 4.31: Graphs illustrate means of DWS from each assessment technique for health profiles 6 to 10

<u>Exploring relationship between DWS from each technique (Correlation Analysis)</u>

Pearson correlation (parametric) and Spearman's Rho (non-parametric) were calculated in order to explore linear relationships between DWS-conditional and DWS-holistic, and between DWS-unconditional and DWS-holistic, for each health profile. Table 4.34 shows the coefficients for both Pearson correlation and Spearman's Rho. The results show that in 5 of the 10 health profiles, profiles #1, 2, 5, 6, and 7, DWS-conditional has higher correlation coefficient with DWS-holistic than DWS-unconditional, indicating that DWS-conditional has stronger association with DWS-holistic for those profiles.

Table 4.34: Pearson correlation and Spearman's Rho coefficients between DWS-conditional and DWS-holistic, and between DWS-unconditional and DWS-holistic

	Pearson Correlation		**Spearman's Rho**	
	Conditional vs Holistic	**Unconditional vs Holistic**	**Conditional vs Holistic**	**Unconditional vs Holistic**
Profile#1	0.134	0.053	0.195	0.103
Profile#2	0.114	0.032	0.104	0.031
Profile#3	0.355	0.332	0.257	0.283
Profile#4	-0.106	0.141	-0.098	0.122
Profile#5	0.242	0.031	0.247	0.094
Profile#6	0.104	0.042	0.105	0.074
Profile#7	0.13	0.027	0.156	0.08
Profile#8	0.108	0.202	0.117	0.216
Profile#9	-0.13	0.264	-0.144	0.27
Profile#10	0.005	0.085	-0.032	0.106

*The grey highlight indicates profiles that have higher coefficients in DWS-conditional than in DWS-unconditional

Comparison of DWSs Across the Three Assessment Techniques

Friedman tests (non-parametric) and GLM-RM (parametric) to test significant differences of DWS across assessment techniques were performed. Results from Friedman tests and GLM-RM are shown in Table 4.35. Note that for all health profiles but profiles 4, 7, and 8, normality assumption for parametric test was not met. Thus, results for those health profiles should refer to the results from non-parametric tests. From Table 4.35, it can be seen that there is no significant difference in DWS among the three assessment techniques for health profiles 1, 4, and 10. For health profiles 2, 6, and 7, DWS-holistic is found to be significantly higher than DWS-conditional and DWS-unconditional. The opposite finding is found for health profiles 3, 5, 8, and 9, in which DWS-holistic is found to be significantly lower than DWS-conditional and DWS-unconditional. Regarding the comparison between DWS-conditional and DWS-unconditional, profiles 2, 3, and 7 have significantly higher DWS-unconditional than DWS-conditional, while for profiles 5, 6, and 9, DWS-unconditional is significantly lower than DWS-conditional. Figure 4.32 and Figure 4.33 present bar graphs for mean DWS for each assessment technique by each profile along with the test results.

Table 4.35: Results from Friedman Tests and GLM-RM for testing significant difference of DWS among the three techniques by each health profile (H = DWS-holistic, UC = DWS-unconditional, and C = DWS-conditional)

Health Profiles	**Friedman Tests (Non-parametric)**			**GLM-RM (Parametric)**		
	Test Statistic (χ^2)	**p-value**	**Post Hoc Tests (by Wilcoxon Signed Rank Tests)**	**Test Statistic (F)**	**p-value**	**Post Hoc Tests (by Bonferroni adjustments)**
Profile 1	0.161	0.923		1.14	0.296	
Profile 2	80.276	<0.001	H > UC H > C UC > C	143.483	<0.001	H > UC H > C UC > C
Profile 3	113.954	<0.001	H < UC H < C UC > C	261.031	<0.001	H < UC H < C UC > C
Profile 4	1.103	0.576		2.078	0.151	
Profile 5	108.621	<0.001	H < UC H < C UC < C	95.58	<0.001	H < UC H < C UC < C
Profile 6	57.379	<0.001	H > UC H > C UC < C	68.605	<0.001	H > UC H > C
Profile 7	11.793	0.003	H > UC H > C UC > C	9.906	0.001	H > UC H > C UC > C
Profile 8	8.759	0.013	H < UC H < C UC < C	7.578	0.006	H < UC H < C
Profile 9	79.966	<0.001	H < UC H < C UC < C	149.143	<0.001	H < UC H < C UC < C
Profile 10	1.402	0.496		1.559	0.217	

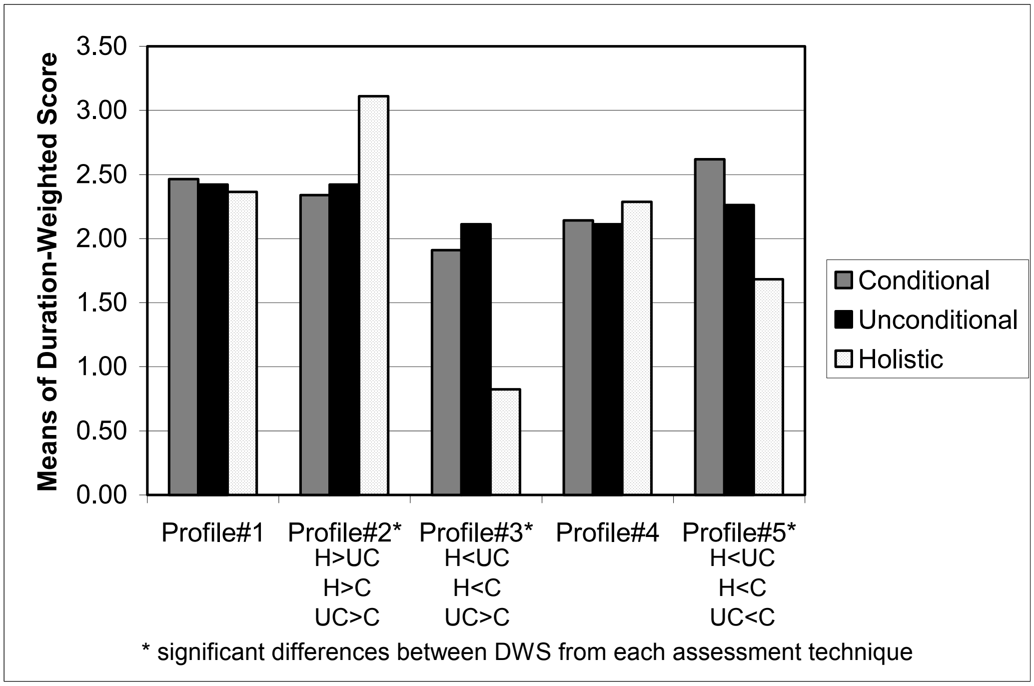


Figure 4.32: Means of DWS across the assessment techniques for each health profile (Profiles 1 to 5)

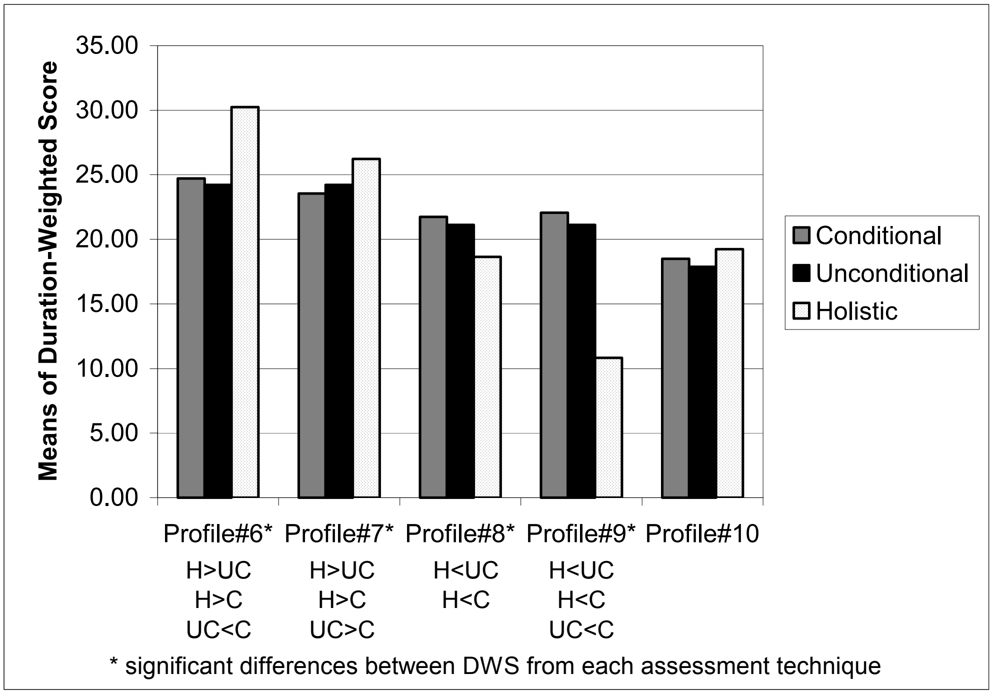


Figure 4.33: Means of DWS across the assessment techniques for each health profile (Profiles 6 to 10)

Comparison of order of preference between health profiles captured by each assessment technique

The results from GLM-RM, Friedman tests, and correlation analysis do not show a great promise that conditional preference assessment predicts holistic preference scores for health profiles better than unconditional preference assessment. However, another approach, which looks at the pattern of the preferences across the health profiles and then compares the significant pattern between each assessment technique was pursued. Since 9 of the 10 DWS for conditional and unconditional assessments are normally distributed, significant differences of DWS across health profiles are identified by running GLM-RM, which also generates pairwise comparisons across health profiles by using a Bonferroni adjustment. However, only DWS-holistic from 5 health profiles are normally distributed. Thus, non-parametric analysis (Wilcoxon Signed Rank test) is also used for pairwise comparisons of DWS-holistic across health profiles. Table 4.36 and Table 4.37 summarize pairwise comparisons along with absolute mean difference of DWS across health profiles #1 to 5 and health profiles #6 to 10, respectively. Figure 4.34 and Figure 4.35 present bar graphs for mean DWS across health profiles for each assessment technique along with the test results for profiles #1 to 5, and profiles #6 to 10, respectively. The following section interprets and explains the results for each assessment technique.

Table 4.36: Pairwise comparison and absolute mean difference of DWS between health profiles (Profiles#1 to 5)

		Pairwise Comparison Between Health Profiles									
		1 vs 2	**1 vs 3**	**1 vs 4**	**1 vs 5**	**2 vs 3**	**2 vs 4**	**2 vs 5**	**3 vs 4**	**3 vs 5**	**4 vs 5**
Conditional	**Std. Error**	0.032	0.030	0.035	0.032	0.042	0.029	0.032	0.042	0.034	0.038
	p-value	0.002	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
	Significance	1 > 2	1 > 3	1 > 4	1 < 5	2 > 3	2 > 4	2 < 5	3 < 4	3 < 5	4 < 5
	Absolute mean difference	0.127	0.555	0.322	0.153	0.429	0.196	0.280	0.233	0.708	0.475
Unconditional	**Std. Error**	0.000	0.005	0.005	0.005	0.005	0.005	0.005	0.000	0.004	0.004
	p-value	.	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	.	<0.001	<0.001
	Significance	1 = 2	1 > 3	1 > 4	1 > 5	2 > 3	2 > 4	2 > 5	3 =4	3 < 5	4 < 5
	Absolute mean difference	0.000	0.309	0.309	0.159	0.309	0.309	0.159	0.000	0.150	0.150
Holistic	**Z**	-6.133	-8.050	-0.571	-6.160	-8.008	-5.993	-7.688	-7.140	-6.742	-3.591
	p-value	<0.001	<0.001	0.568	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	0.002
	Significance	1 < 2	1 > 3		1 > 5	2 > 3	2 > 4	2 > 5	3 < 4	3 < 5	4 > 5
	Absolute mean difference	0.746	1.541	0.078	0.681	2.287	0.824	1.427	1.463	0.860	0.603

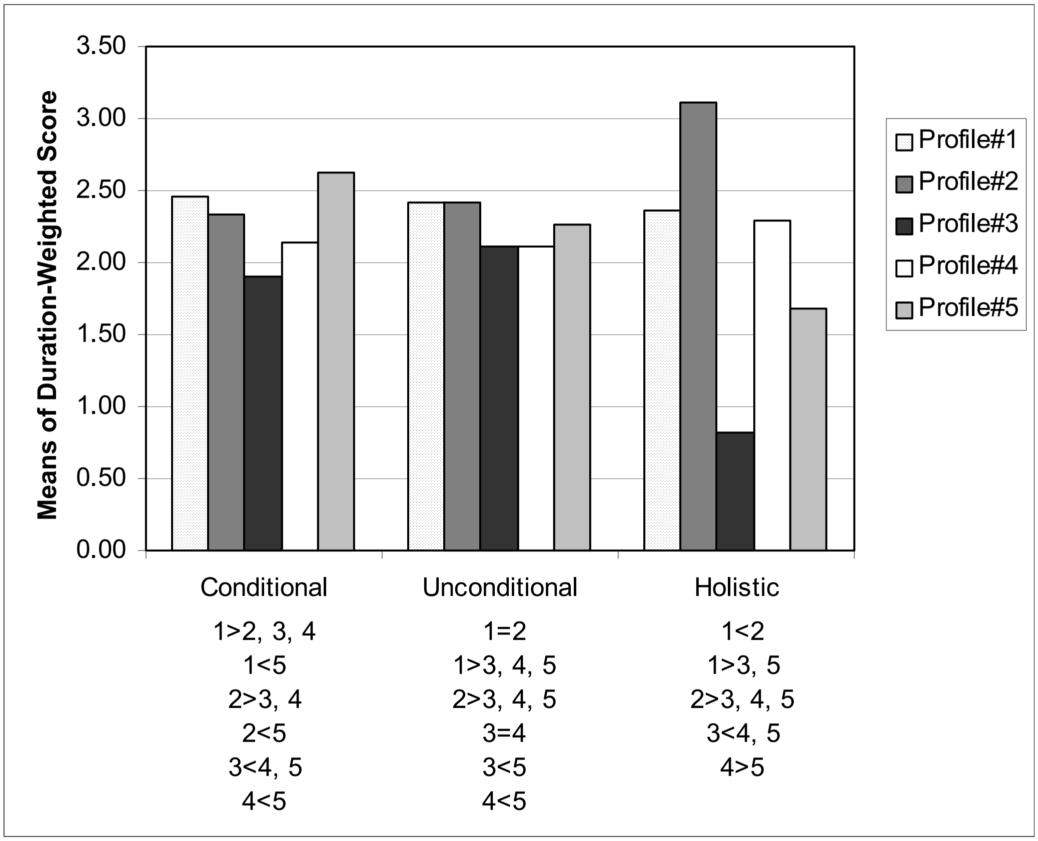


Figure 4.34: Means of DWS and the significant results across health profiles (Profiles #1 to 5) for each assessment technique

Table 4.37: Pairwise comparison and absolute mean difference of DWS between health profiles (Profiles#6 to 10)

		Pairwise Comparison Between Health Profiles									
		6 vs 7	6 vs 8	6 vs 9	6 vs 10	7 vs 8	7 vs 9	7 vs 10	8 vs 9	8 vs 10	9 vs 10
Conditional	**Std. Error**	0.325	0.239	0.327	0.318	0.358	0.271	0.281	0.396	0.345	0.241
	p-value	0.005	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	1.000	<0.001	<0.001
	Significance	6 > 7	6 > 8	6 > 9	6 > 10	7 > 8	7 > 9	7 > 10		8 > 10	9 > 10
	Absolute mean difference	1.179	2.978	2.665	6.220	1.799	1.486	5.041	0.313	3.242	3.555
Unconditional	**Std. Error**	0.000	0.049	0.049	0.052	0.049	0.049	0.052	0.000	0.052	0.052
	p-value	.	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	.	<0.001	<0.001
	Significance	6 = 7	6 > 8	6 > 9	6 > 10	7 > 8	7 > 9	7 > 10	8 = 9	8 > 10	9 > 10
	Absolute mean difference	0.000	3.087	3.087	6.348	3.087	3.087	6.348	0.000	3.261	3.261
Holistic	**Z**	-3.875	-6.947	-7.974	-6.784	-4.573	-8.058	-5.635	-5.769	-0.552	-6.004
	p-value	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	0.581	<0.001
	Significance	6 > 7	6 > 8	6 > 9	6 > 10	7 > 8	7 > 9	7 > 10	8 > 9		9 > 10
	Absolute mean difference	4.026	11.597	19.396	11.014	7.571	15.370	6.988	7.799	0.583	8.383

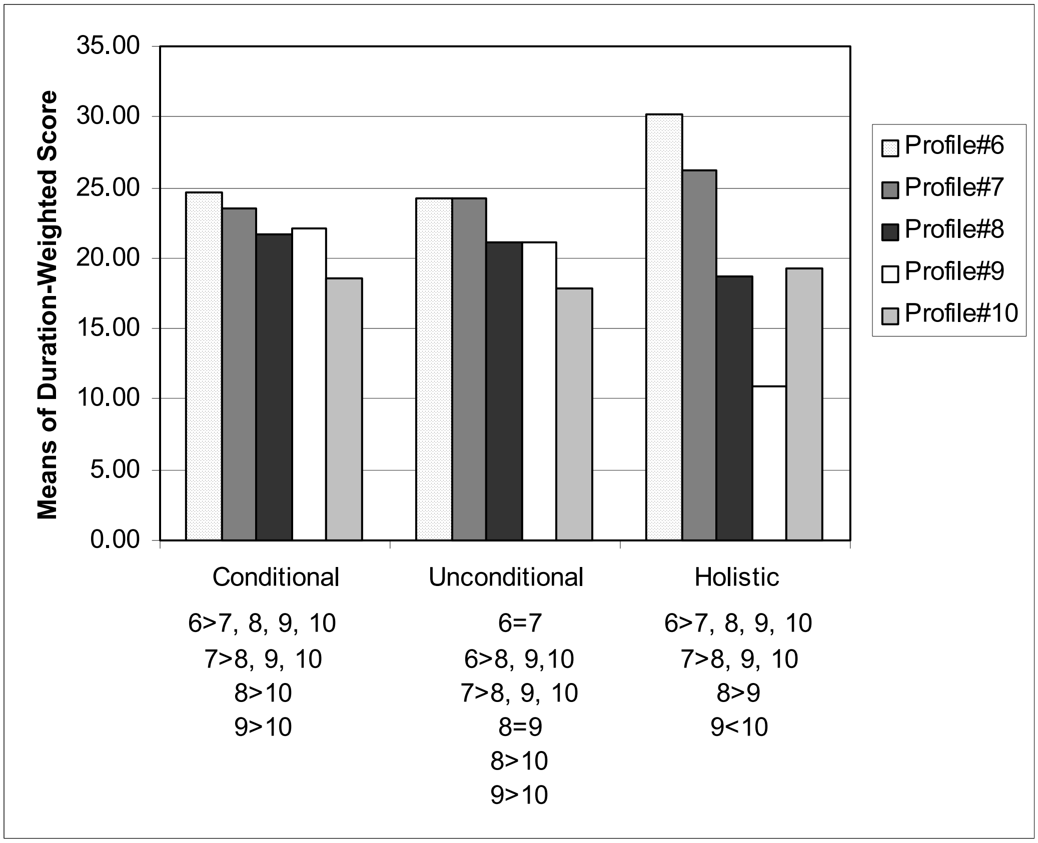


Figure 4.35: Means of DWS and the significant results across health profiles (Profiles#6 to 10) for each assessment technique

Conditional Preference Assessment Technique

For 4-year health profiles (profiles #1 to 5), Table 4.36 and Figure 4.34 demonstrate that DWS for health profile #1 is significantly higher than DWS for health profiles #2, 3, and 4 by 0.127, 0.555, and 0.322, respectively. DWS for health profile #2 is significantly higher than DWS for health profiles #3 and 4 by 0.429 and 0.196. Health profile #5 has significantly larger DWS than health profiles #1, 2, 3 and 4 with mean differences of 0.153, 0.280, 0.708, and 0.475, correspondingly. Furthermore, DWS for health profile #4 is significantly larger than DWS for health profile #3 by 0.233.

Regarding lifetime health profiles (profiles #6 to 10), Table 4.37 and Figure 4.35 show that DWS for health profile #6 is significantly higher than DWS for health profiles #7, 8, 9, and 10 by 1.179, 2.978, 2.665, and 6.220, respectively. DWS for health profile #7 is significantly larger than DWS for health profiles #8, 9, and 10 by 1.799, 1.486, and 5.041, respectively. Additionally, DWS for health profile #10 is significantly smaller than DWS for health profiles #8 and 9 by 3.242 and 3.555, correspondingly.

Unconditional Preference Assessment Technique

For 4-year health profiles (profiles #1 to 5), Table 4.36 and Figure 4.34 show that DWS for health profile #1 is equal to DWS for health profile#2 which is significantly higher than DWS for health profiles #3, 4, and 5 by 0.309, 0.309, and 0.159, respectively. Furthermore, DWS for health profile #3 is equal to DWS for health profile #4, which is significantly less than DWS for health profile #5 by 0.150.

Regarding lifetime health profiles (profiles #6 to 10), Table 4.37 and Figure 4.35 show that DWS for health profile #6 is equal to DWS for health profile #7 which is significantly higher than DWS for health profiles #8, 9, and 10 by 3.087, 3.087, and 6.348, respectively. Additionally, DWS for health profile #8 is equal to DWS for health profile #9, which is significantly larger than DWS for health profile #10 by 3.261.

By design, the similarity of DWS for health profiles #1 and 2, health profiles #3 and 4, health profiles #6 and 7, and health profiles #8 and 9 corresponds to the fact that each pair consists of exactly the same four health states but in different patterns. With the unconditional preference assessment method, those pairs of profiles are assigned the same DWS since the pattern of the health profile does not influence the unconditional assessment.

Holistic Preference Assessment Technique

For 4-year health profiles (profiles #1 to 5), Table 4.36 and Figure 4.34 show that DWS for health profile #2 is significantly higher than DWS for health profiles #1, 3, 4 and 5 by 0.746, 2.287, 0.824, and 1.427, respectively. DWS for health profile #1 is significantly larger than DWS for health profiles #3 and 5 by 1.541 and 0.681. Health profile #3 has significantly smaller DWS than health profiles #4 and 5 with mean difference of 1.463, and 0.860, correspondingly. Furthermore, DWS for health profile #4 is significantly larger than DWS for health profile #5 by 0.603.

Regarding lifetime health profiles (profiles #6 to 10), Table 4.37 and Figure 4.35 show that DWS for health profile #6 is significantly higher than DWS for health profiles #7, 8, 9, and 10 by 4.026, 11.597, 19.396, and 11.014, respectively. DWS for health profile #7 is significantly larger than DWS for health profiles #8, 9, and 10 by 7.571, 15.370, and 6.988, respectively. Additionally, DWS for health profile #9 is significantly smaller than DWS for health profiles #8 and 10 by 7.799 and 8.383, correspondingly.

From the results described above, it can be seen that there are similarities and differences among the three techniques in paired comparisons of profiles. Table 4.38 summarizes the inconsistent findings across assessment techniques. Conditional preference assessment shows a closer similarity to holistic preference assessment than unconditional preference assessment in predicting order of preference between health profiles #3 and 4, and between health profiles #6 and 7. On the other hand, unconditional preference assessment shows a better prediction in predicting order of preference (as compared to order of preference predicted by holistic preference assessment) between health profiles #1 and 5, and between health profiles #2 and 5. Notice that DWS-unconditional between health profiles #1 and 2, profiles #3 and 4, profiles #6 and 7, profiles #8 and 9, are equal for each pair since the two profiles in each pair contain the exact same health states but different in sequences. Unconditional preference assessment assumes additive independence and thus the sequence of health states does not affect the DWS-unconditional preference score. Violation of the additive independence assumption has been shown (Spencer, 2003; Mackeigan et al., 1999; Kupperman et al., 1997; Richardson et al., 1996). Thus, unconditional preference assessment fails to capture order of preference between health profiles that consist of the same health states but are different in sequences. Further analyses in these health profiles were performed in order to explore if the conditional preference assessment can capture preference between pairs of health profiles that the unconditional preference assessment cannot capture.

Table 4.38: Summary of the differences in significant trends of preference between assessment techniques (NS = Not significant)

Comparison between health profiles#	**Holistic**	**Conditional**	**Unconditional**
1 and 2	2 > 1	1 > 2	1 = 2
1 and 4	NS	1 > 4	1 > 4
1 and 5	1 > 5	1 < 5	1 > 5
2 and 5	2 > 5	2 < 5	2 > 5
3 and 4	3 < 4	3 < 4	3 = 4
4 and 5	4 > 5	4 < 5	4 < 5
6 and 7	6 > 7	6 > 7	6 =7
8 and 9	8 > 9	NS	8 = 9
8 and 10	NS	8 > 10	8 > 10
9 and 10	9 < 10	9 > 10	9 > 10

Analysis of preference between pairs of health profiles (health profiles # 1-2, 3-4, 6-7, and 8-9)

Analyses at this stage were performed in order to explore if conditional preference assessment can capture the preferences between pairs of health profiles that unconditional preference assessment fails to capture. Two approaches were taken:

Compare the choice predicted from conditional preference score assessment with the actual choice from paired comparison

As described earlier, in Phase 2 experiment, the subjects performed paired-comparisons for all possible combinations of the five profiles within each group (4-year profiles group and lifetime profiles group) with the aim of generating the ranking among the five profiles. In this analysis, for each subject, the preferred profiles assessed from conditional preference assessment were identified by the higher DWS-conditional between each health profile pairs of interest. Then the preferred profiles identified were compared to the preferred profiles obtained from the paired comparisons. Table 4.39 shows the number and the percentage of subjects in which the preferred profiles from conditional preference assessment match those from the paired comparisons.

Table 4.39: Numbers and percentage of subjects in which the preferred profiles from conditional preference assessment match those from the paired comparisons

Comparison between health profiles#	Number of matched cases (Total N = 87)	Percentage of matched cases
1 and 2	31	36%
3 and 4	61	70%
6 and 7	49	56%
8 and 9	38	44%

From Table 4.39, the highest percentage of matched cases occurs at the comparison between health profiles #3 (steeply decreasing 4-year profile) and 4 (steeply increasing 4-year profile). For 70% of the 87 subjects, the preferred profile between health profiles #3 and 4 assessed by conditional preference assessment is consistent with the preferred profile selected by the subjects as a result of a paired comparison. On the other hand, 56% of matched cases are found for comparison between health profiles #6 (gradually decreasing lifetime profile) and 7 (gradually increasing lifetime profile), 44% for comparison between health profiles #8 (steeply decreasing lifetime profile) and 9 (steeply increasing lifetime profile), and 36% for comparison between health profiles #1 (gradually decreasing 4-year profile) and 2 (gradually increasing 4-year profile).

Compare the choice predicted from conditional preference score assessment with the choice predicted from holistic preference scores

For each subject, the preferred profiles assessed from conditional preference assessment were identified by the higher DWS-conditional between each health profile pairs of interest. Moreover, the preferred profiles assessed from holistic preference assessment were identified by the higher DWS-holistic between each pair of health profile. Then the preferred profiles identified from conditional preference assessment were compared to those identified from holistic preference assessment. Table 4.40 shows the number of subjects and the percentage in which the preferred profiles from conditional preference assessment match those from the holistic preference assessment.

Table 4.40: Numbers and percentage of subjects in which the preferred profiles from conditional preference assessment match those from holistic preference assessment

Comparison between health profiles#	**Number of matched cases (Total N = 87)**	**Percentage of matched cases**
1 and 2	39	45%
3 and 4	62	71%
6 and 7	44	51%
8 and 9	36	41%

From Table 4.40, the highest percentage of matched cases occurs at the comparison between health profiles #3 (steeply decreasing 4-year profile) and 4 (steeply increasing 4-year profile). For 71% of the 87 subjects, the preferred profile between health profiles #3 and 4 assessed by conditional preference assessment is consistent with a preferred profile assessed by holistic preference assessment. On the other hand, the number drops to 51% of matched cases for comparison between health profiles #6 (gradually decreasing lifetime profile) and 7 (gradually increasing lifetime profile), 45% for comparison between health profiles #1 (gradually decreasing 4-year profile) and 2 (gradually increasing 4-

year profile), and 41% for comparison between health profiles #8 (steeply decreasing lifetime profile) and 9 (steeply increasing lifetime profile).

Compare the significant effects of the three factors (duration, slope, direction) across techniques

As described earlier, 8 of the 10 health profiles in Phase 2 experiment were constructed by varying three factors in a 2^3 factorial design. Those three factors are direction (2 levels: improvement or deterioration over time), duration (2 levels: 4 years and lifetime), and rate of change in health status over time (2 levels: gradual and steep). Another two health profiles have no systematic pattern, one for each duration level. Thus, it is worthy to investigate the effects of the three factors that are captured by each assessment technique. Analyses in this step were carried out by GLM-RM by treating duration, slope, and direction as independent variables with DWS for each assessment technique as a dependent variable. The results are shown in Table 4.41.

Table 4.41: Results from GLM-RM for testing effects of duration, slope, direction, and their interactions on DWS from each assessment technique

Factors	**Conditional**		**Unconditional**		**Holistic**	
	F-Statistic	p-value	F-Statistic	p-value	F-Statistic	p-value
Duration	18482.53	<0.001	65658.24	<0.001	2206.096	<0.001
Slope	237.017	<0.001	3894.01	<0.001	251.957	<0.001
Direction	1.383	0.243			25.396	<0.001
Duration x Slope	145.628		3894.01	<0.001	214.898	<0.001
Duration x Direction	2.897	0.092			54.504	<0.001
Slope x Direction	19.963	<0.001			7.283	0.008
Duration x Slope x Direction	8.726	0.004			15.942	<0.001

*The grey highlight indicates significant effect at alpha = 0.05

The results from Table 4.41 show that, as expected, all the main and interaction effects of duration, slope, and direction are significant for DWS-holistic, with p-values up to 0.008. For DWS-conditional, all effects but the direction main effect and the 2-way interaction effect between duration and direction are significant, with p-values ranged from less than 0.001 to 0.004. However, at alpha = 0.10, the 2-way interaction effect between duration and direction will be significant (p-value = 0.092). For DWS-unconditional, direction or any interaction effect of direction with other factors are not applicable since direction or sequence of health states has no effect on DWS-unconditional due to the additive independence assumption. Duration, slope, and their interaction are significant with p-values less than 0.001. This also provides some support to the fact that the DWS-conditional method performs better than DWS-unconditional.

Summary of the Results

In Phase 2 experiment, different analyses were run with the aim of testing whether the conditional preference score assessment can better predict the preference for the entire health profile than the unconditional preference score assessment. Table 4.42 summarizes the analyses that have been carried out and their results.

Table 4.42 Summary of the results in Phase 2

Types of Analyses	Conclusions
Correlation analysis	DWS-conditional is more highly correlated with DWS-holistic than DWS-unconditional in 5 of the 10 health profiles.
Comparison of DWS across the three assessment techniques for each health profile	There are no significant differences across DWS from the three assessment techniques in three of the 10 health profiles. For the remaining 7 profiles, both DWS-conditional and DWS-unconditional are significantly different from DWS-holistic. The results do not clearly demonstrate that conditional preference score assessment can better predict holistic preference score than unconditional preference score assessment in any health profile.
Comparison of order of preference between health profiles captured by each assessment technique	Eight of the 20 possible paired comparisons of preference between DWS-conditional across health profiles are not consistent with the order of preference predicted by DWS-holistic and 7 of the 20 possible comparisons of order of preference between DWS-unconditional are not consistent with the order of preference from DWS-holistic. The results do not clearly demonstrate that conditional preference score assessment can better predict holistic preference score than unconditional preference score assessment.
Analysis of preference between pairs of health profiles (health profiles #1-2, 3-4, 6-7, and 8-9)	While unconditional preference score assessment cannot predict the preferred choice between these pairs of health profiles due to additive independence assumption, conditional preference score assessment can predict the preferred choice correctly in approximately 70% of cases. The results suggest that conditional preference score assessment performs well when unconditional preference assessment fails by design due to violation of the additive independence.
Comparison of the significant effects of the three factors (duration, slope, direction) across assessment techniques	Conditional preference score assessment performs better than the unconditional preference score assessment in that the significant results of the three factors are more similar to the results from the holistic preference score assessment. This is due to the fact that the unconditional preference score assessment cannot predict the significant effect of “direction” because of its additive independence assumption.

CHAPTER 5

CONCLUSION AND DISCUSSION

5.1 Conclusion and Discussion

In this study, a conditional preference assessment technique, a new method to capture preferences for multistate health profiles was developed and tested. The experiments were designed and conducted to investigate its validity. A total of six hypotheses were formulated and examined in two phases of the study.

In the first phase of the study, the potential effect of the relationship between current health state and future health state on preference judgments of the future health state was explored. Subjects were presented with different hypothetical health scenarios. Each scenario was composed of two different health states: a current health state and a future health state. Various scenarios were created by varying direction of change between health states, amplitude of change, duration of the current health state, and level of the future health state. For each scenario, a conditional preference score for future health state was elicited using a direct rating through a visual analog scale. Five of the total of six hypotheses were investigated in this phase of the study. The hypotheses involved the significance of the main effects and interactions of the factors of interest: direction of change between health states, amplitude of change, and duration of the current health state. The results were analyzed separately for each level of the future health state.

As shown in Table 4.29, the testing of all hypotheses in Phase 1 resulted in the following findings:

- The three-way interaction effect between direction of change, amplitude and current health state duration on the conditional preference score of the future health state is significant only in the design with a higher level of future health state (Design A), but not in the other two lower levels of future health state (Design B and C).
- The two-way interaction effect between direction and amplitude of change from current health state to future health state is significant in all three designs.
- The two-way interaction effect between direction of change and current health state duration is significant only in Design A, but not in Design B and C.
- The two-way interaction effect between amplitude of change from current health state to future health state and current health state duration is significant only in Design A, but not in Design B and C.
- An effect of current health state duration resulted in different relative impact of direction of change and amplitude of change across the designs.

As expected, it was found that characteristics of the current health state have an effect on preference judgments for future health states. However, unexpectedly, the nature of the effects varied across different levels of the future health state, indicating that in addition to the fact that preference for a future health state depends on current health state, the nature and extent of the impact of the current health state characteristics in assessing preference for a future health state (i.e. direction of change, amplitude of change, and current health state duration) also depend on the level of the future health state itself.

For health scenarios at a higher level of future health state (Design A – future health state level at 0.70), all main effects of direction, amplitude, and duration, and all the combinations of their interactions are significant. Regarding health scenarios at a lower level of future health state (Design C – future health state level at 0.40), only the main effects of direction and amplitude, and their interaction are significant. For health scenarios at a medium level of future health state (Design B – future health state level at 0.55), only the interaction effect between direction and amplitude is significant. These results imply that preferences for future health state are more sensitive to the current health state as the future health state is better.

In order to better explain and understand the findings regarding the interdependence between current health state and future health state on the preference scores, Figure 5.1, Figure 5.2, and Figure 5.3 graphically illustrate the average conditional preference scores across the subjects for scenarios in Design A, B, and C (the average scores were presented numerically in Tables 4.5, 4.7, 4.9, and 4.11 in Chapter 4). Additional analysis, the Wilcoxon Signed Ranks test, was performed in order to test the difference between the mean conditional preference score and the mean actual future health state score assigned by the subjects, for each scenario. Table 5.1 summarizes the discrepancies and the test results. (Note that the positive value of discrepancy indicates that the actual future health state score is higher than its conditional preference score, and vice versa).

Table 5.1: Descriptive statistics of the discrepancy between the actual future health state score and the conditional preference score along with the results from Wilcoxon Signed Ranks test

Scenario	Mean of the Discrepancy	S.D.	Wilcoxon Signed Ranks Test	
			Z	p-value
Scenario 1-A	0.276	0.184	-0.627	0.531
Scenario 2-A	0.023	0.167	-7.982	<0.001
Scenario 3-A	0.013	0.115	-3.08	0.002
Scenario 4-A	0.016	0.133	-2.786	0.005
Scenario 5-A	0.041	0.143	-0.263	0.793
Scenario 6-A	0.042	0.158	-0.938	0.348
Scenario 7-A	0.042	0.122	-2.917	0.004
Scenario 8-A	0.042	0.139	-2.461	0.014
Scenario 1-B	-0.023	0.128	-0.906	0.365
Scenario 2-B	-0.017	0.138	-1.822	0.068
Scenario 3-B	-0.013	0.128	-0.863	0.388
Scenario 4-B	-0.012	0.127		0.006
Scenario 5-B	0.042	0.157	-1.239	0.215
Scenario 6-B	0.004	0.158	-1.265	0.206
Scenario 7-B	-0.012	0.123	-0.783	0.433
Scenario 8-B	-0.016	0.133	-0.565	0.572
Scenario 1-C	-0.039	0.131	-3.649	<0.001
Scenario 2-C	-0.057	0.140	-2.601	0.009
Scenario 3-C	-0.047	0.129	-1.141	0.254
Scenario 4-C	-0.049	0.142	-2.975	0.003
Scenario 5-C	-0.077	0.216	-3.263	0.001
Scenario 6-C		0.215	-3.255	0.001
Scenario 7-C	-0.029	0.145	-2.087	0.037
Scenario 8-C	-0.043	0.153	-4.846	<0.001

* The grey highlight indicates significant difference at alpha = 0.05

By looking at Table 5.1 along with Figure 5.1, Figure 5.2, and Figure 5.3, it can be clearly seen that conditional preference scores are significantly different from the actual future health state scores, indicating that current health state affects the judgment of the future health state, almost in every scenario in Design A and Design C, but not in Design B. Another interesting finding is that the mean conditional preference score in Design A is always *lower* than the actual future health state score (positive values of the discrepancies). On the other hand, in Design C, conditional preference score is always *higher* than the actual future health state score (negative values of the discrepancies). In Design B, there is virtually no significant difference between conditional preference score and actual future state score. Indeed, in Design B, the conditional preference score is approximately equal to the actual future health state score, as shown by the results from the analysis. This overall finding is important: independently of direction of change, amplitude of change and duration of current health state, scoring a future health state with information about the current health state affects the future health state score downward if the future health state is high and upward if the future health state is low. These findings signify that the pattern of the relationship between current health state and the future health state on the judgment of the future health state depends upon the level of the future health state itself.

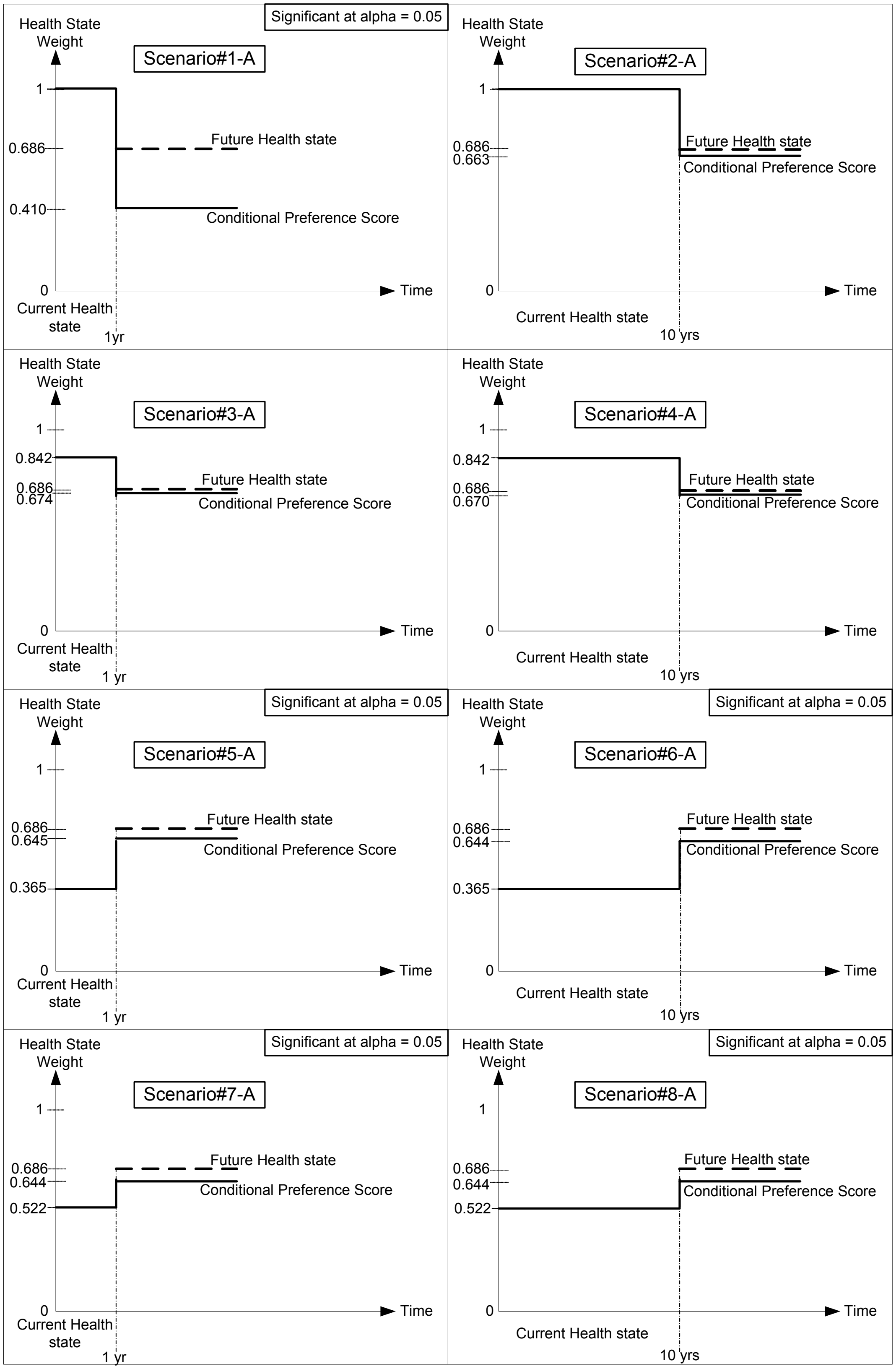


Figure 5.1: Graphical illustration of means of conditional preference scores for Design A

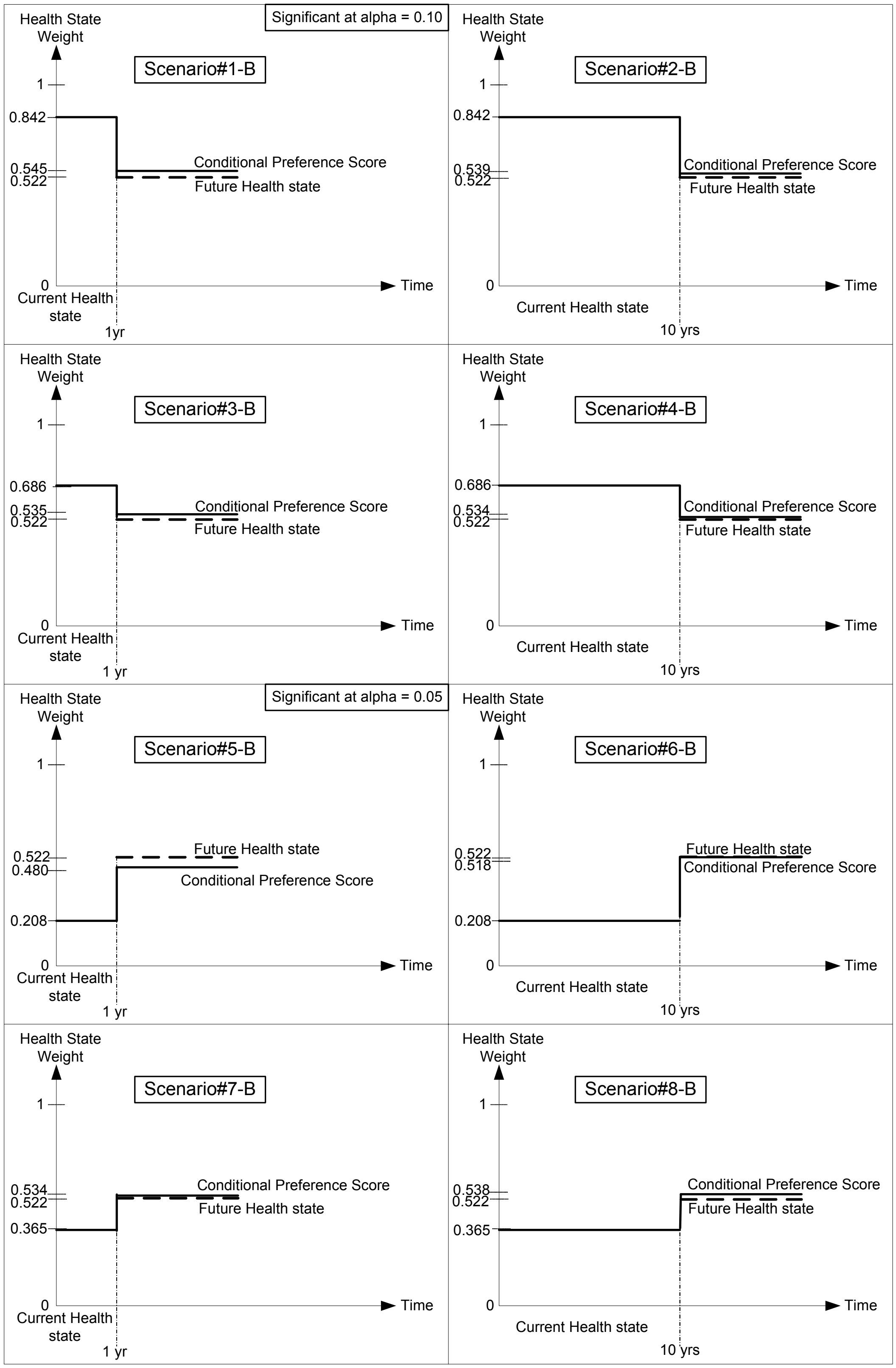


Figure 5.2: Graphical illustration of means of conditional preference scores for Design B

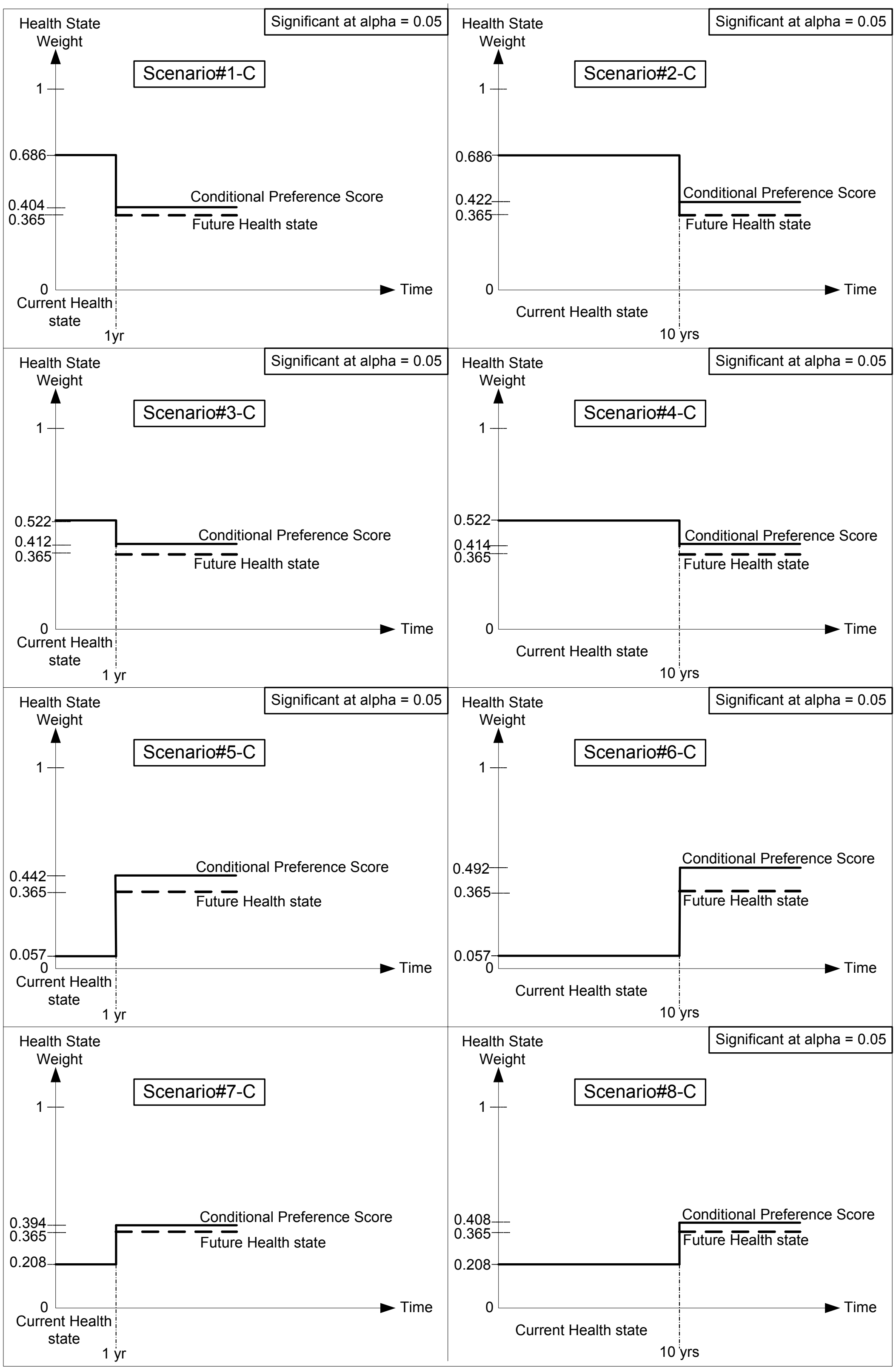


Figure 5.3: Graphical illustration of means of conditional preference scores for Design C

The hypotheses set in this phase anticipated a significant effect of 2-way interaction between duration and amplitude and a smaller effect of current health state on the preference score of future health state when current health state duration is short (1 year) as compared to when the current health state duration is long (10 years).

Regarding the expected significant effect of the 2-way interaction between duration and amplitude, larger differences between conditional preference scores and the actual future health state scores were expected for scenarios that have high amplitude of change from current health state to future health state (scenarios 1, 2, 5, and 6) than for scenarios that have low amplitude (scenarios 3, 4, 7, and 8). This is only found for the decreasing scenarios in Design A (scenarios 1-A and 2-A versus scenarios 3-A and 4-A) and the increasing scenarios in Design C (scenarios 5-C and 6-C versus scenarios 7-C and 8-C). However, the magnitude of difference is not substantial. Additionally, lower conditional preference scores as compared to the actual future health state level are expected for decreasing scenarios (scenarios 1 to 4) and higher conditional preference scores as compared to the actual future health state level are expected for increasing scenarios (scenarios 6 to 8). For Design A (see Figure 5.1), decreasing scenarios (scenarios 1-A to 4-A) have lower conditional preference scores than the actual future health state scores as expected; however, none of the increasing scenarios (scenarios 5-A to 8-A) have higher conditional preference scores than the actual future health state scores, conflicting to the expectation. For Design B and Design C (see Figure 5.2 and Figure 5.3, respectively), none of the decreasing scenarios (scenarios 1-B to 4-B and scenarios 1-C to 4-C) have lower conditional preference scores than the actual future health state scores, inconsistent to the expectation. However, increasing scenarios in

Design C (scenarios 6-C to 8-C) have higher conditional preference scores than the actual future health state scores which are as expected.

The findings described above demonstrate that the results behave in a way similar to what was expected regarding the interaction effect between direction and amplitude but only for decreasing scenarios in Design A and increasing scenarios in Design C. A plausible explanation is that individuals experiencing a high level of health state (up to perfect health) as in scenarios 1-A to 4-A, tend to be more sensitive when they anticipate losing their perfect (or close to perfect) health, as markedly shown by the results in scenario 1-A. Thus, such individuals tend to undervalue the scores for their future health state when comparing to a high reference point, a close to perfect current health state. However, when considering medium levels of current health states, individuals who have been staying in poorer health may not see additional deterioration of their health as a significant issue due to, potentially, being psychologically adjusted to a poorer health state to start with. On the other hand, when they are in health state that is very low (close to death), as in scenario 6-C to 8-C, individuals, realizing that their health is going to be improved, seem to over-evaluate the future (improved) health state. As a result, they are likely to score their future improved health state higher than what they would have done in the absence of knowledge of a very poor current health state. However, when individuals are not being in very poor health, the improvement in their health may not be seen as a significant change for them. As can be seen in Design B, starting at a medium health state level, changing their health state in either direction does not seem to make any difference in individuals' perceptions and scoring regarding their future health. Thus the results strongly support that preferences for future health states strongly and systematically depend on where the current health state is. Moreover, the pattern of the interdependence also depends upon the level of future health state itself.

Concerning the effect of duration, larger differences between conditional preference scores and actual future health states were expected for scenarios with longer current health state duration (scenarios 2, 4, 6, and 8) as compared to scenarios with shorter current health state duration (scenarios 1, 3, 5, and 9). In examining Figure 5.1, Figure 5.2, and Figure 5.3, none of the designs overall shows any such trend. However, when looking at the analysis for each duration separately, interesting factors of the current health state (direction and amplitude) play a more significant role when the current health state duration is short than when the current health state duration is long (but only for Design A and B, high and medium future health state levels respectively). While this finding lends support to rejecting the hypotheses, a plausible explanation is that individuals, when staying in the current health state in a short period of time, do not psychologically adjust to the health state and thus, judgments for any changes in their health state in the future is affected by the current health state. On the other hand, when individuals have been staying in their current health state for a long period of time, they tend to adjust themselves and do not necessarily relate their judgment on their current health state. However, this does not hold for Design C (low level of future health state), since the results in Design C show that all characteristics of current health state are significant independently of current health state duration. No matter what duration in Design C, conditional preference scores for future health states are higher than the unconditional future health state scores. It only appears that the discrepancy between conditional and unconditional score is slightly higher as the duration of the current health state is shorter.

In the second phase of the study, the aim was to investigate whether the proposed decomposition technique (conditional preference assessment) can better predict preference score for a full health profile than the conventional decomposition technique (unconditional preference assessment). Thus, in addition to the experiment in the first phase, subjects were presented with different hypothetical health profiles and were asked to assess holistic preference scores for each health profile by providing a direct rating through visual analog scale. Different health profiles were created by varying direction of the profiles, rate of change and duration of the profiles. Each health profile was composed of four different health states. Duration-weighted scores were calculated for scores obtained from conditional, and unconditional preference assessments and were compared to holistic scores and to each other.

Unfortunately, the results from the analyses do not show great promise that the new proposed decomposition technique in which several conditional preference score assessments are carried out and integrated to produce a preference score for the entire health profile predicts holistic preference scores for health profiles significantly better than the decomposition technique that uses unconditional preference score assessments. Duration-weighted scores from conditional and unconditional preference score assessments were similar to each other while the holistic preference scores were different from them especially in some health profiles. However, when looking at the holistic preference assessment itself, it should be recognized that the assessment tasks themselves can be problematic. Several studies as well as this study have proposed other decomposition techniques in order to aid with the difficulties that arise in formulating holistic preference score assessment.

When making judgment on the entire multistate health profile, holistic preference assessment can be difficult and complicated to individuals, especially when individuals are making judgment on hypothetical health scenarios that are far away from their current health experience. In this and other studies, young healthy subjects asked to make judgment of their hypothetical future health scenarios may have difficulties, especially with health scenarios that are very distant from their current (generally) healthy state. Take, for example, health profile# 9 which started with a health state that was slightly better than death (e.g. state 33333 – confined to bed, unable to wash and dress yourself, unable to perform your usual activities, extreme pain or discomfort, and extremely anxious or depressed) and therefore is extremely different from what their current health state is. For that scenario, the average holistic preference score obtained is remarkably low (as compared to the average score obtained from health profile #8 which is composed of the exact same four health states but started with perfect health) although health profile #9 has a rapid improvement to perfect health in the last 10 years of their life expectancy. Another important element to consider in interpreting holistic judgments is that individuals tend to use their expectations as a tool in making holistic preference assessment, as suggested by Chapman (1996a). People judge their preference for the health profile by examining how close it is to their expectations. The closer the profile is to their expectation, the higher the preference for that health profile tends to be. For example, the relatively low score obtained from health profile #9 as previously pointed out can be explained by the fact that this health profile, which has an improvement over time when individuals are getting older, is not consistent with potential expectations. As can be seen in the results for lifetime health profiles, decreasing sequences tend to be preferred to increasing sequences. On the other hand, regarding the 4-year health profiles, the results show that increasing sequences are preferred to decreasing sequences. These findings are again consistent with the findings from Chapman (1996a). However, it is still questionable if holistic preference scores assessed, regardless of whichever strategies individuals use, are fully reliable in terms of representing their preferences regarding the hypothetical health profiles.

As previously mentioned regarding the results in Phase 2 experiment, duration-weighted scores from conditional and unconditional preference assessments are similar to each other; nonetheless, they both are quite different from holistic preference scores. One explanation of the deviation of holistic scores from both conditional and unconditional duration-weighted scores can be identified as the use of the entire scale versus the integration of component scores. When the subjects make holistic judgment, if they perceive that the health profile is extremely good or bad, they can give the score for that health profile at the very high or low end on the scale. However, with the conditional and unconditional preference score assessment methods, it is less likely to obtain extreme scores on either end of the scale since the scores for the entire health profile come from the integration of several component scores through a weighted additive calculation. In order to obtain extreme scores for the health profile in the case of conditional and unconditional preference score assessment, all or most of the component scores that come from individual health state assessments have to be consistently high or low, which is not possible for the profiles constructed in this study. The health profiles used in this study are composed of several health states that ranged from very poor health to perfect health; thus, it is less likely that the subjects would assign extremely low score for a perfect health or extremely high score for a very poor health state. Thus, as shown in Figure 4.30 and Figure 4.31, conditional and unconditional scores across all profiles display a more horizontal and flatter profiles than the holistic scores do.

Although the analyses do not show that conditional preference scores predict holistic preference scores better than unconditional preference scores, some of the results indicate that conditional preference assessment can predict preferred choices between pairs of health profiles better than unconditional preference assessment. This is important since, while individuals have difficulties in rating health profiles (i.e., producing a holistic score), they fare much better in selecting preferred profiles in forced paired comparisons, as designed in this study. In the case that unconditional preference assessment cannot predict trend of preference due to the additive independence assumption, for example, profiles that are composed of the exact same components of health states but different in sequences, unconditional preference assessment would automatically produce identical duration weighted scores across the profiles, indicating that those profiles are equally preferred. In contrast, conditional preference assessments in which the interdependence relationship between a pair of health states is taken into account, were generally, although not all the time, indicative of the actual direction of preference in this case. Thus, from the perspective of paired comparisons across health profiles, the conditional method performed better than the unconditional method. Moreover, the results from Phase 1 experiment strongly support the fact that an interdependence relationship between current health state and future health state exists and matters in making preference assessments. Thus, it is essential for further research to perform future studies in order to develop a method that is able to integrate the interdependence relationship between health states into the preference assessment technique in the case of multistate health profiles. The simple method proposed here, while a good first step did not completely live up to expectations in terms of solving the problem.

Regarding the experiment performed in this study, it should be noted that it was very extensive and demanding from a cognitive processing perspective, especially because of the use of EQ-5D. In particular, this is the case of the health profiles in the study composed of four EQ-5D health states. In the paired comparisons, it was not easy to compare a profile with four consecutive EQ-5D health states with another profile that had another four consecutive EQ-5D health states. Moreover, the number of assessments in this study was quite large: 20 unconditional preference score assessments, 31 conditional preference score assessment, 20 paired comparisons between health profiles, and 10 holistic preference score assessments. The subject population for this study consisted mostly of students from the School of Industrial and Systems Engineering at the Georgia Institute of Technology. Dealing with a high-cognitive demand task as in this experiment might not be an issue for this group of subjects. However, implementing the same experimental design with other segments of the population, especially in terms of their educational levels, will be difficult and changes to the experimental settings should be made. The number of assessments should be greatly reduced by using fewer replications or a fractional design. The experimental protocol may require a walkthrough and a face-to-face interview instead of having the subjects complete the tasks through the experimental computer program on their own. Regarding the health state representation, use of multimedia techniques for health state representation, which allows subjects to visualize or experience the hypothetical health scenarios, will make the experiment easier and feasible with a more diverse population.

Another interesting point to be discussed is the representation of the current health state duration, one of the investigated factors in the experiment. One may argue that it was difficult for the subjects to imagine and take the duration factor into account when assessing the hypothetical health scenarios. Thus, for future research, the use of multimedia health state representation, or other techniques, is suggested.

Regarding the design of the experiment, the selection of the 7 health states required in order to construct the health scenarios and the health profiles was based on the 20 unconditional preference scores each subject assigned at the beginning of the experiment. One may question the use of different health states across the subjects to create the final health scenarios in the design of experiment. However, using the same health states in the health scenarios and the health profiles for all subjects almost certainly would have created an issue with respect to differences in the perception regarding each health state and in the ordering of the health states across subjects, which in turn result in the inability to control the amplitude size and the direction of change between current health state and future health state, which were the investigated factors.

Moreover, one may wonder whether starting the current health state at perfect health as opposed to some other health state that would be perceived as relatively high on the judgment scale but lower than perfect health may affect the results. However, there were only two scenarios from the total of 24 scenarios that started current health state at perfect health and moreover only one of them produced significantly lower conditional preference score for the future health state as compared to the actual future health state score when assessed independently. Thus, by replacing the current health state with some other health state that would be perceived as relatively high on the scale (but less than perfect health) instead of using the perfect health, the effect of the current health state factors explored might have been smaller. However, the difference in the findings would not greatly affect the results.

Concerning the implications for other related fields, the results from this study show that current health state and its characteristics have an effect on individuals' judgments regarding their future health states. Thus, practitioners and clinicians, who want to capture patients' preferences regarding their health states in order to make use of them in medical decision making, should definitely ensure that every future health state considered is conditioned on the patients' current health state. Regarding the implications from a societal perspective, specifically cost effectiveness analyses, this study suggests that the effectiveness of interventions designed to improve health for people with very low current health state might be underestimated. The results from this study show that people with very low current health state value their future health state at a higher level. Thus, with the assessment from the existing decomposition technique, the effectiveness of the interventions may be undervalued. This implies that some interventions that are not currently seen as cost effective might in fact be cost effective.

5.2 Limitations of the study

One of the limitations in this study concerns the assessment technique used to capture preferences. A direct rating method, the visual analog scale (VAS), was the assessment technique used throughout this study. While the VAS is attractive because it is simple, easy to administer, and the subjects can perform the task by themselves, the validity of the VAS approach is still questionable. VAS scores represent the strength of the preference under certainty. Thus, VAS is not a utility measurement technique and thus cannot fully represent individuals' preferences under risk and uncertainty. Moreover, some biases may be associated with the use of VAS. For example, the VAS score for a health state also depends on the number of better or worse health states presented previously or at the same time. If many better health states are shown before or simultaneously with the health state assessed, then the score assigned might be undervalued. On the other hand, if several worse health states are shown before or at the same time with the health state assessed, its score might be overvalued. However, even with known issues, VAS has been acceptable and is widely used in similar research and studies for health outcome measurement and health economics. Using the VAS technique, while not a gold standard of utility measurement, allowed us to perform the study and revealed important relationships.

Another limitation involves the use of EQ-5D system to describe health states in the study. Although the description of the health states in five dimensions is easy to understand and does not require the subjects to have knowledge regarding diseases or any specific disease-related health conditions, the description of the five dimensions may not be specific enough and thus the subjects may have difficulty in imagining themselves in those health states and thus may have made additional unknown assumptions regarding the health states being considered and evaluated. Take, for example, the fourth dimension concerning pain or discomfort, subjects need to make an assumption about the types of pain or discomfort or specific symptoms that are the results of that pain or discomfort. Other examples are the second and the third dimensions, which involve ability to wash and dress themselves and ability to perform their usual activities, respectively. These two elements are seen as components that may relate to each other. Thus, it is hard for the subjects to imagine having severe (or some) problem with one dimension but some problem (or no problem) with the other dimension. However, the EQ-5D system remains an acceptable and widely used tool in health outcome measurement today.

5.3 Directions for Future Research

The results from this study confirm that there is a relationship between two consecutive health states that has an effect on the judgment of the future health state. However, this study does not successfully develop a new measurement method that is able to integrate all the significant relationships uncovered into a new calculation of preference scores for evaluating complex health profiles. Thus, future research would be important to be performed in at least three general directions, as described below.

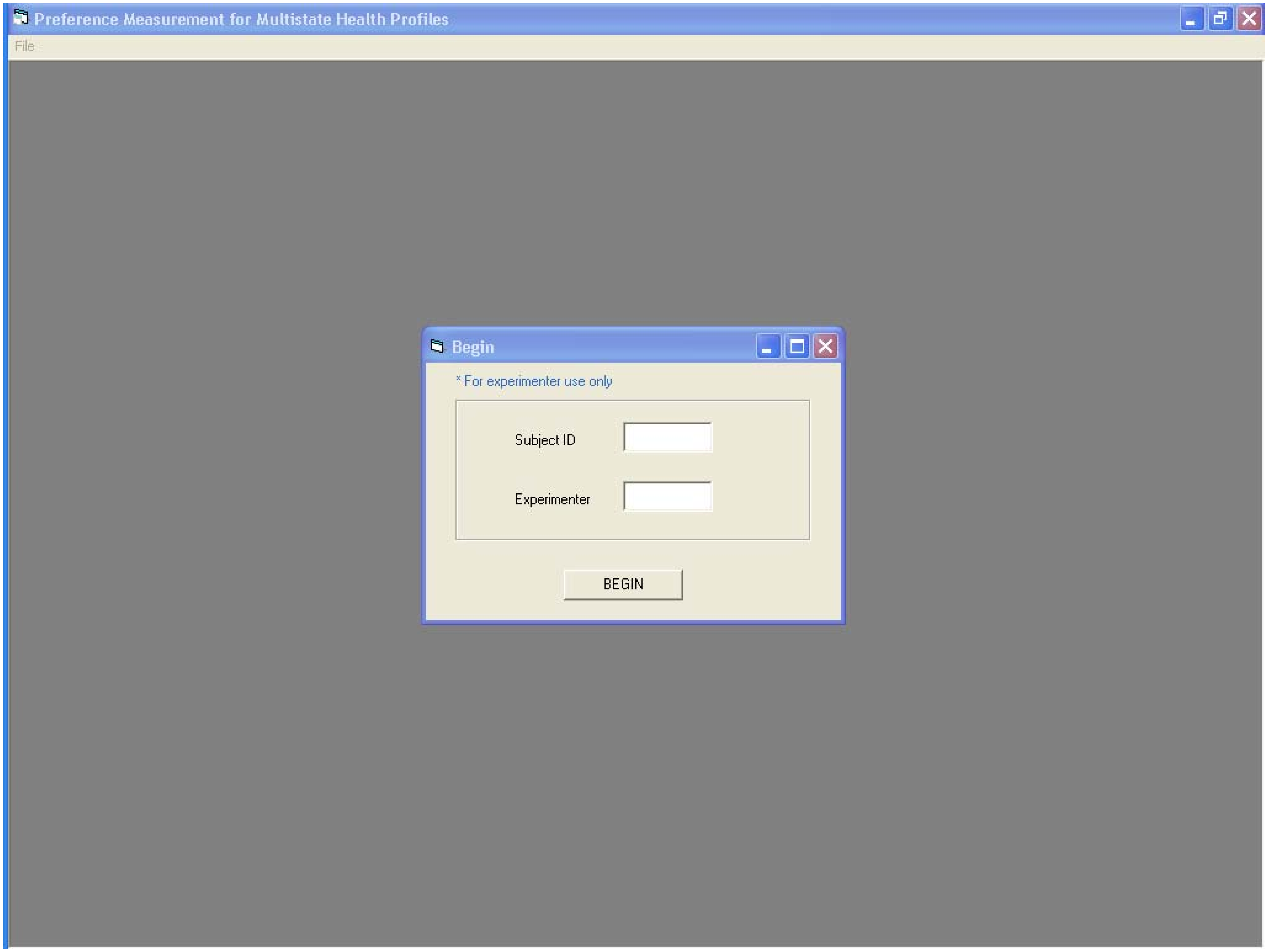
First, future research is in need to further explore the nature and extent of the relationship between health states such as exploring complex relationships between more than two consecutive health states.

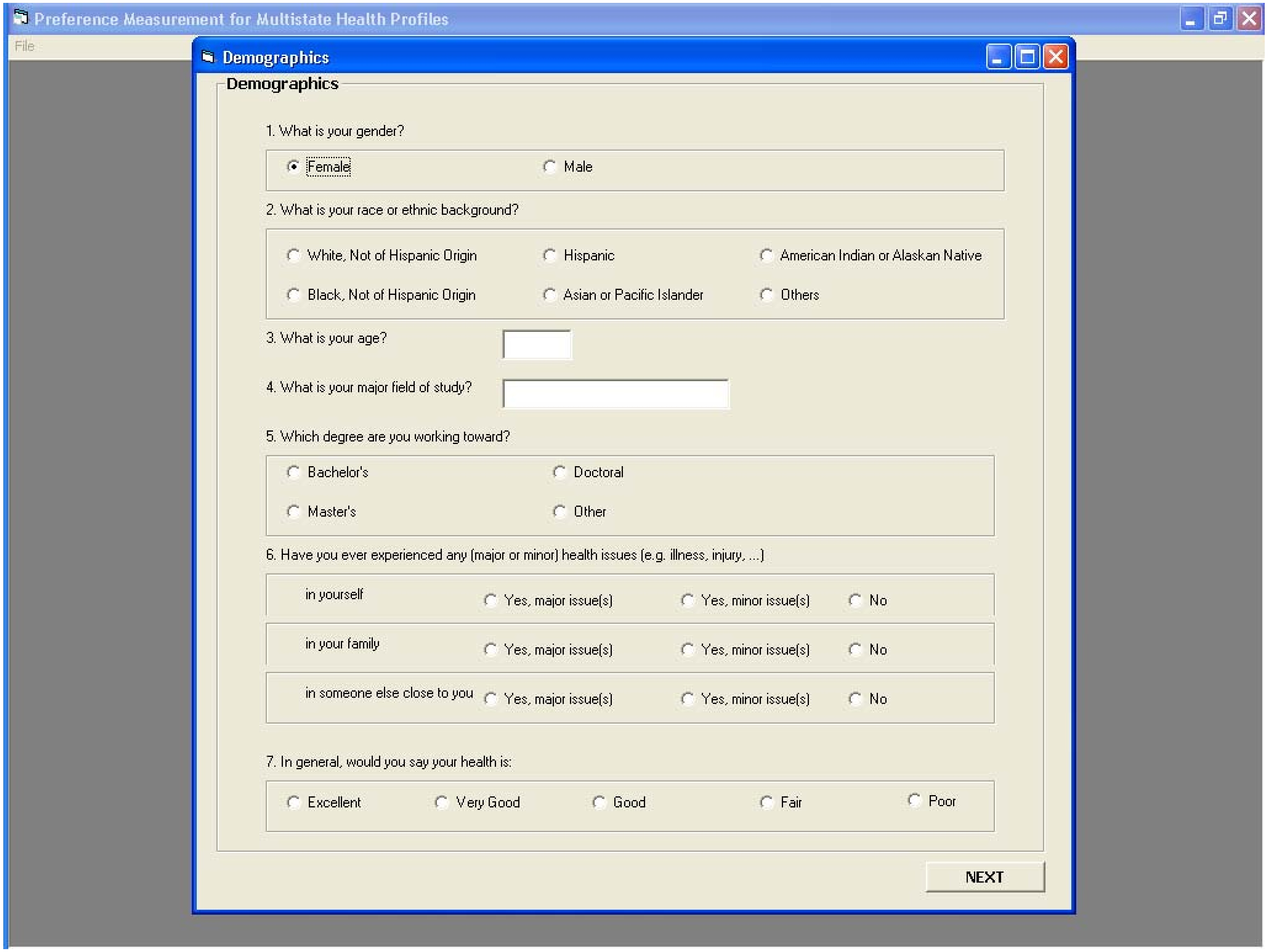
Second, since the visual analog scale may suffer from limitations as described above, the use of other assessment techniques such as time-tradeoff and standard gamble in replicating and expanding the same experiments would also be important to investigate.

Finally, performing similar experimental studies with different subject populations, such as using actual patients (with experienced health states) instead of young healthy subjects (with hypothetical health scenarios) would be another important direction for future research in multistate health profile preference assessment.

APPENDIX A

SCREENSHOTS OF THE COMPUTER PROGRAM USED IN THE EXPERIMENT


Preference Measurement for Multistate Health Profiles
File
Begin
* For experimenter use only
Subject ID
Experimenter
BEGIN



Preference Measurement for Multistate Health Profiles

File

Demographics

Demographics

1. What is your gender?

( ) Female ( ) Male

2. What is your race or ethnic background?

( ) White, Not of Hispanic Origin ( ) Hispanic ( ) American Indian or Alaskan Native

( ) Black, Not of Hispanic Origin ( ) Asian or Pacific Islander ( ) Others

3. What is your age?

4. What is your major field of study?

5. Which degree are you working toward?

( ) Bachelor's ( ) Doctoral

( ) Master's ( ) Other

6. Have you ever experienced any (major or minor) health issues (e.g. illness, injury, ...)

in yourself ( ) Yes, major issue(s) ( ) Yes, minor issue(s) ( ) No

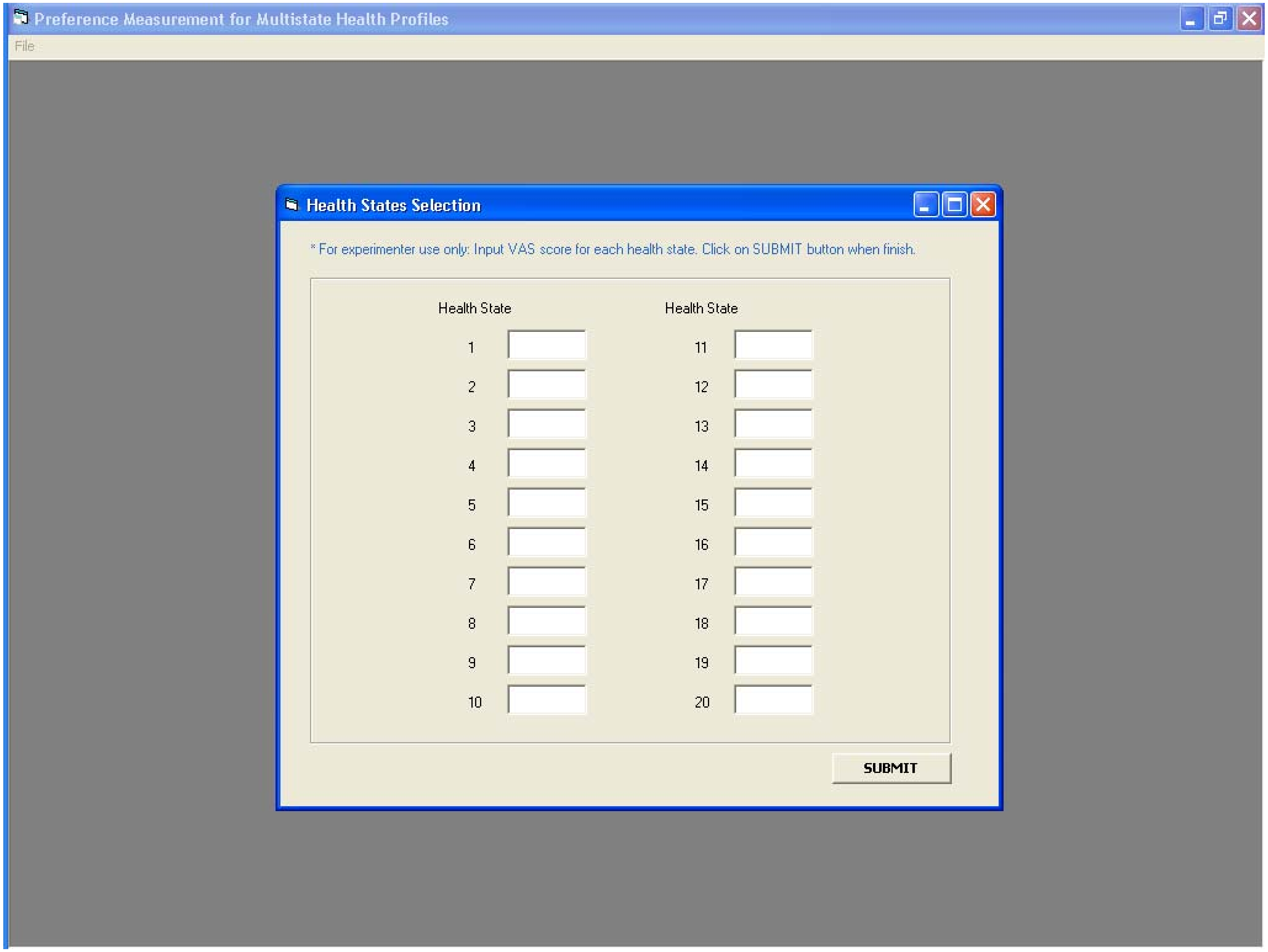
in your family ( ) Yes, major issue(s) ( ) Yes, minor issue(s) ( ) No

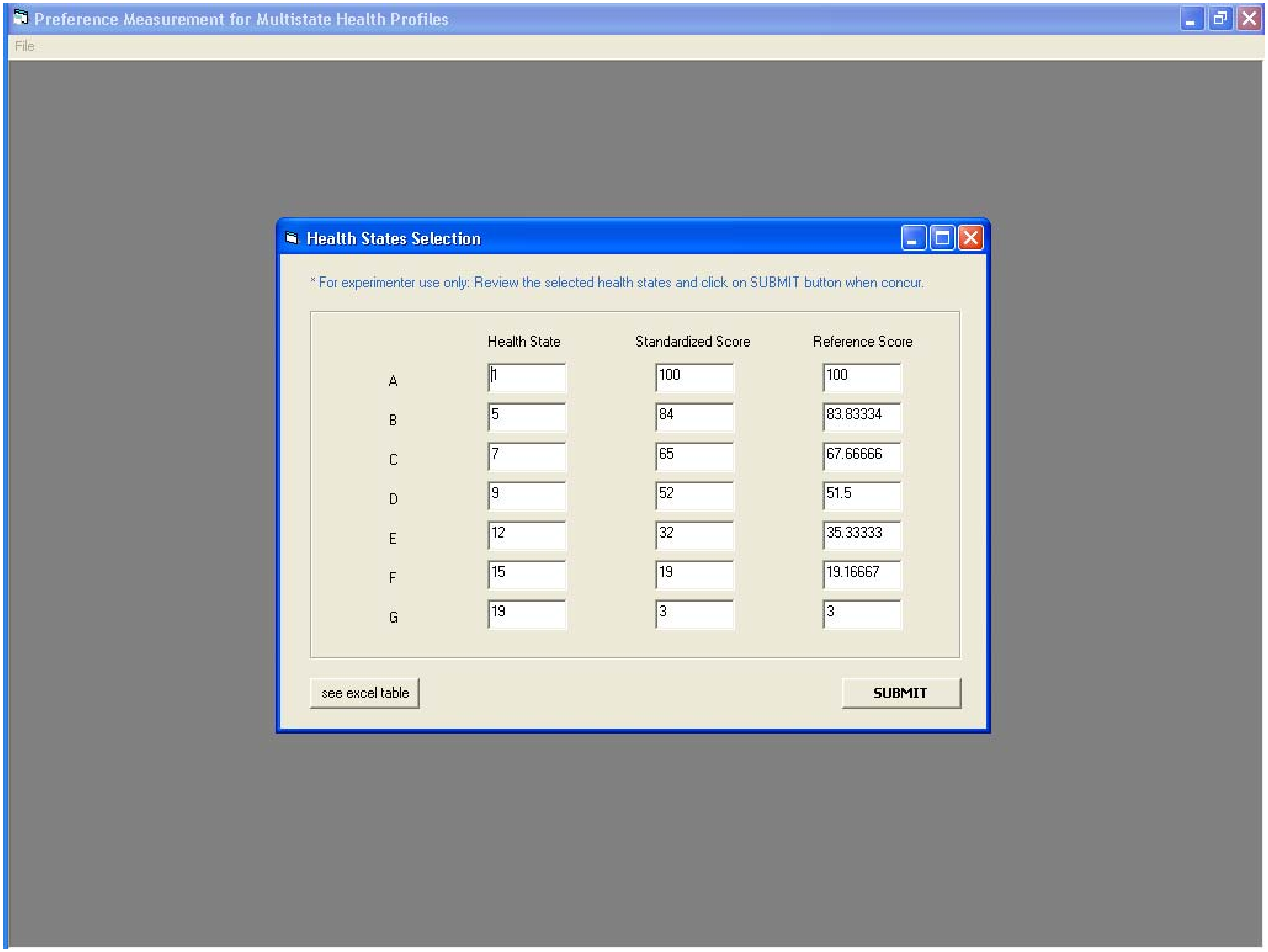
in someone else close to you ( ) Yes, major issue(s) ( ) Yes, minor issue(s) ( ) No

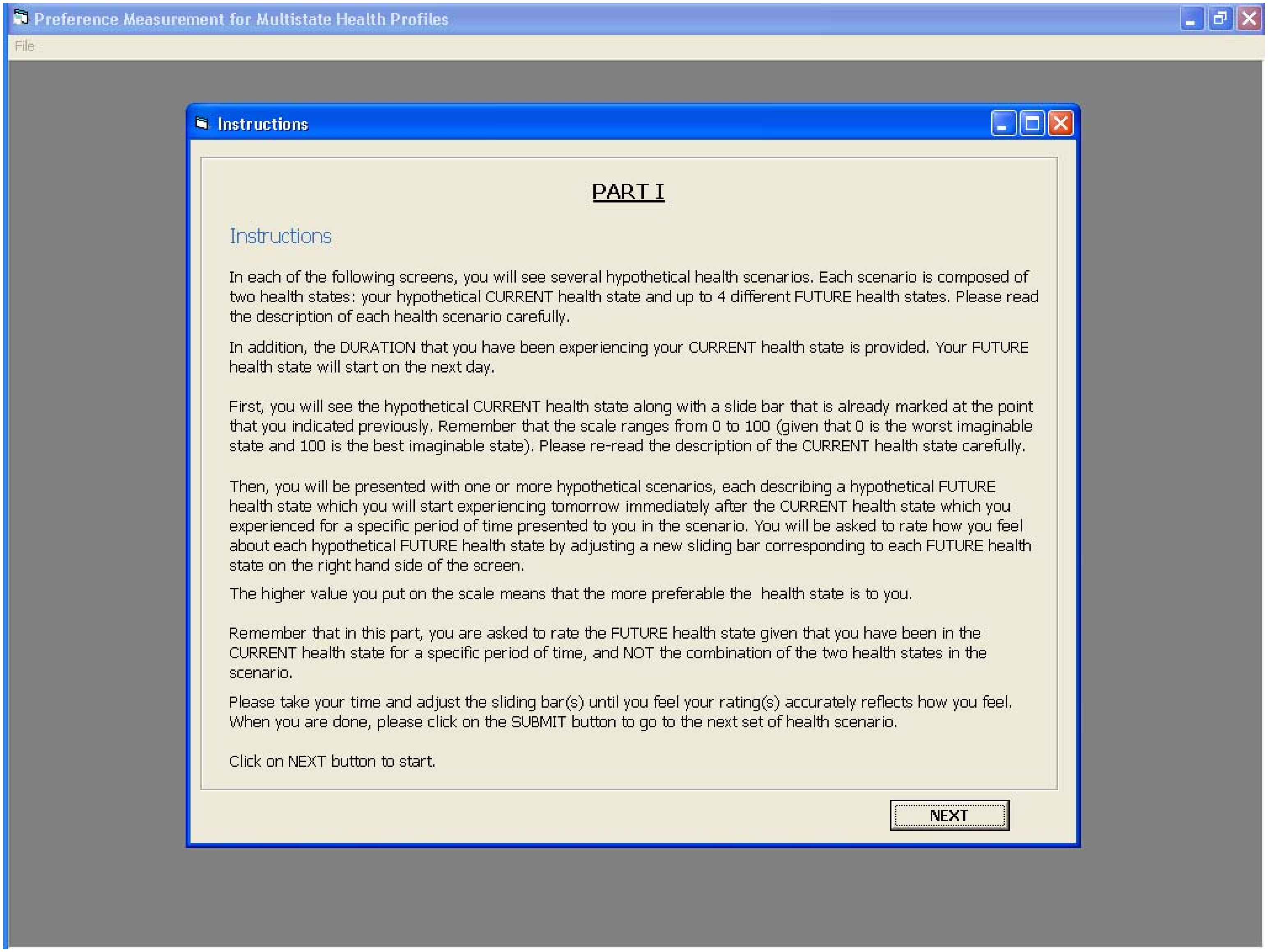
7. In general, would you say your health is:

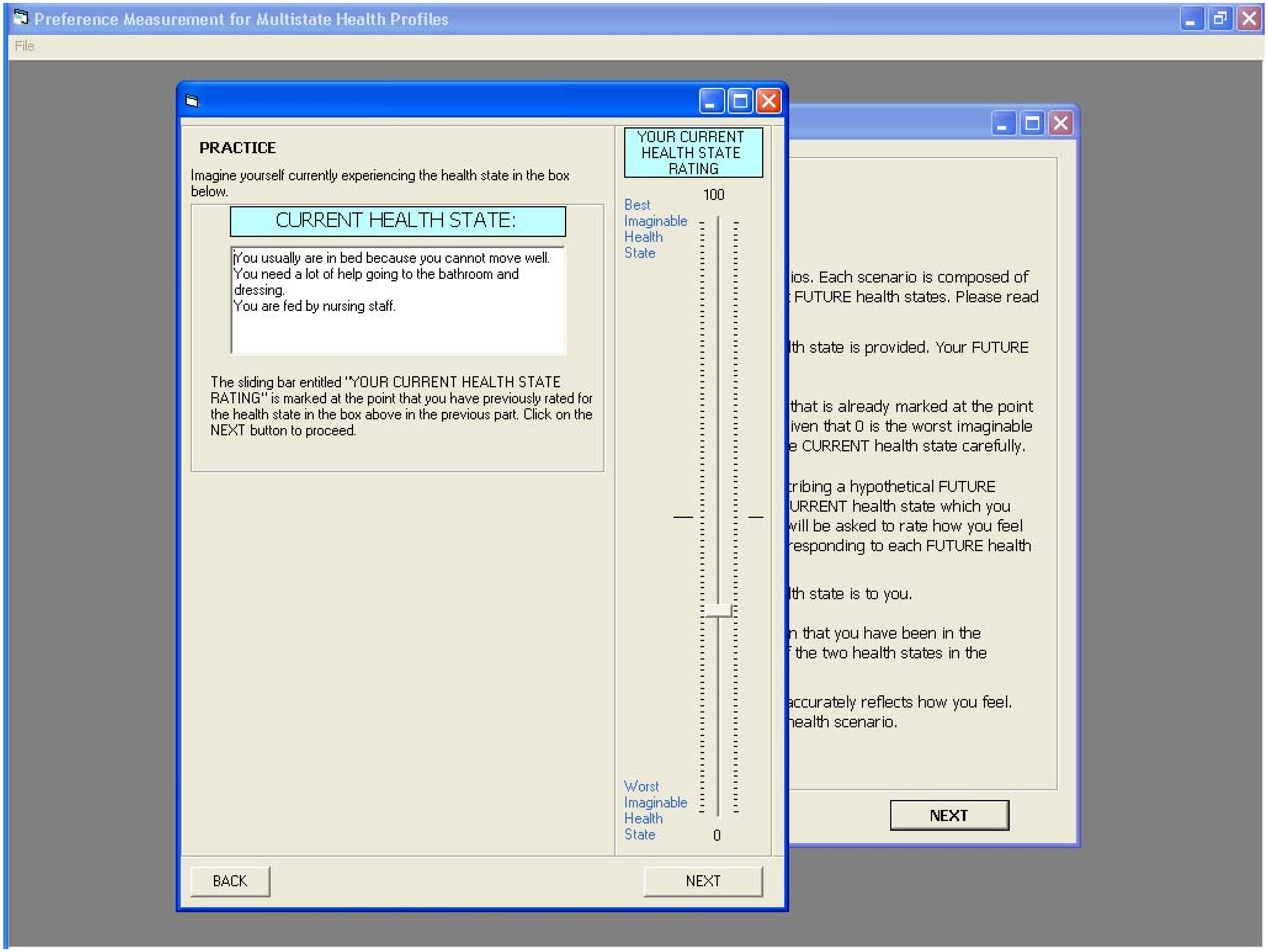
( ) Excellent ( ) Very Good ( ) Good ( ) Fair ( ) Poor

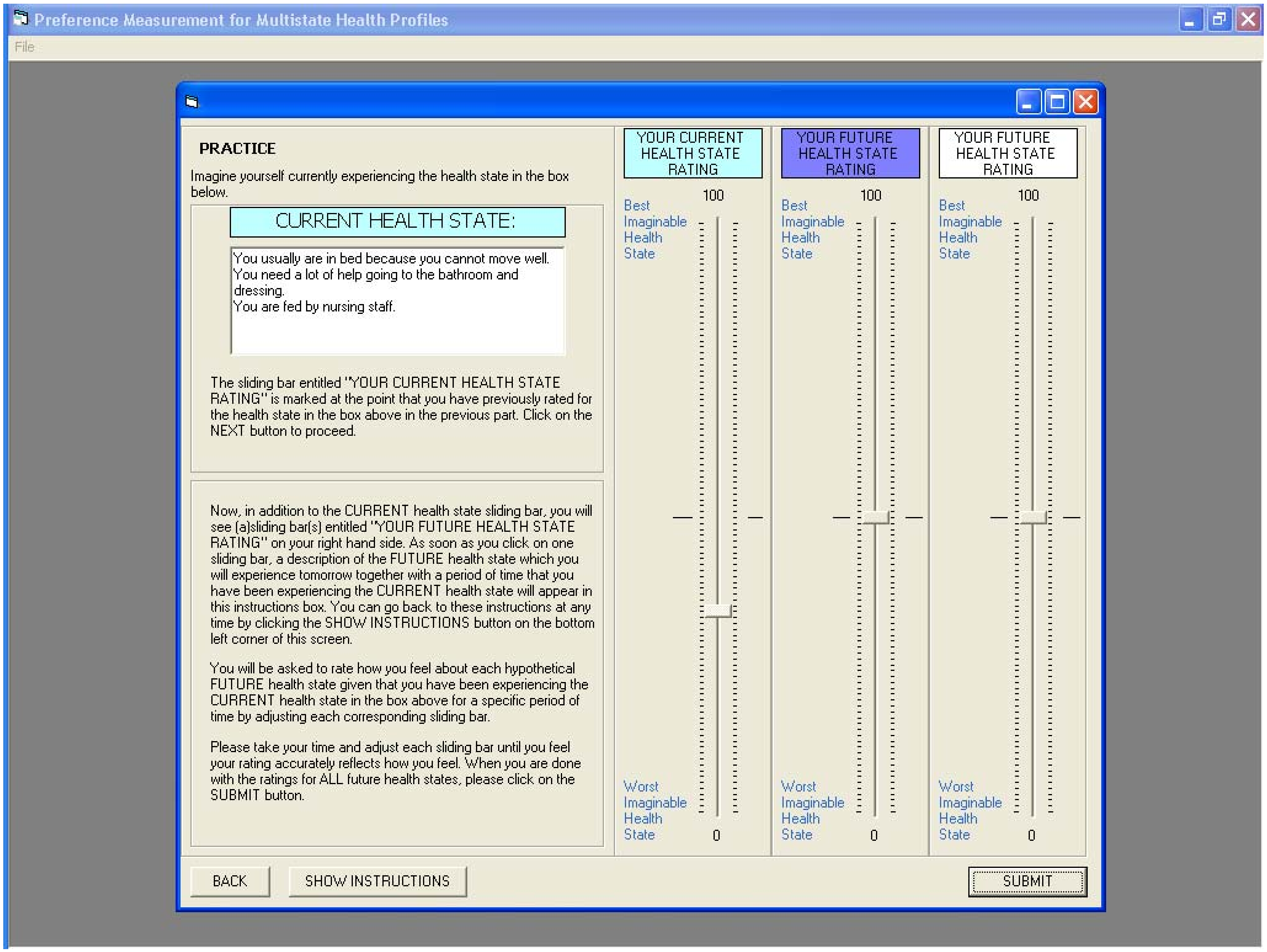
NEXT

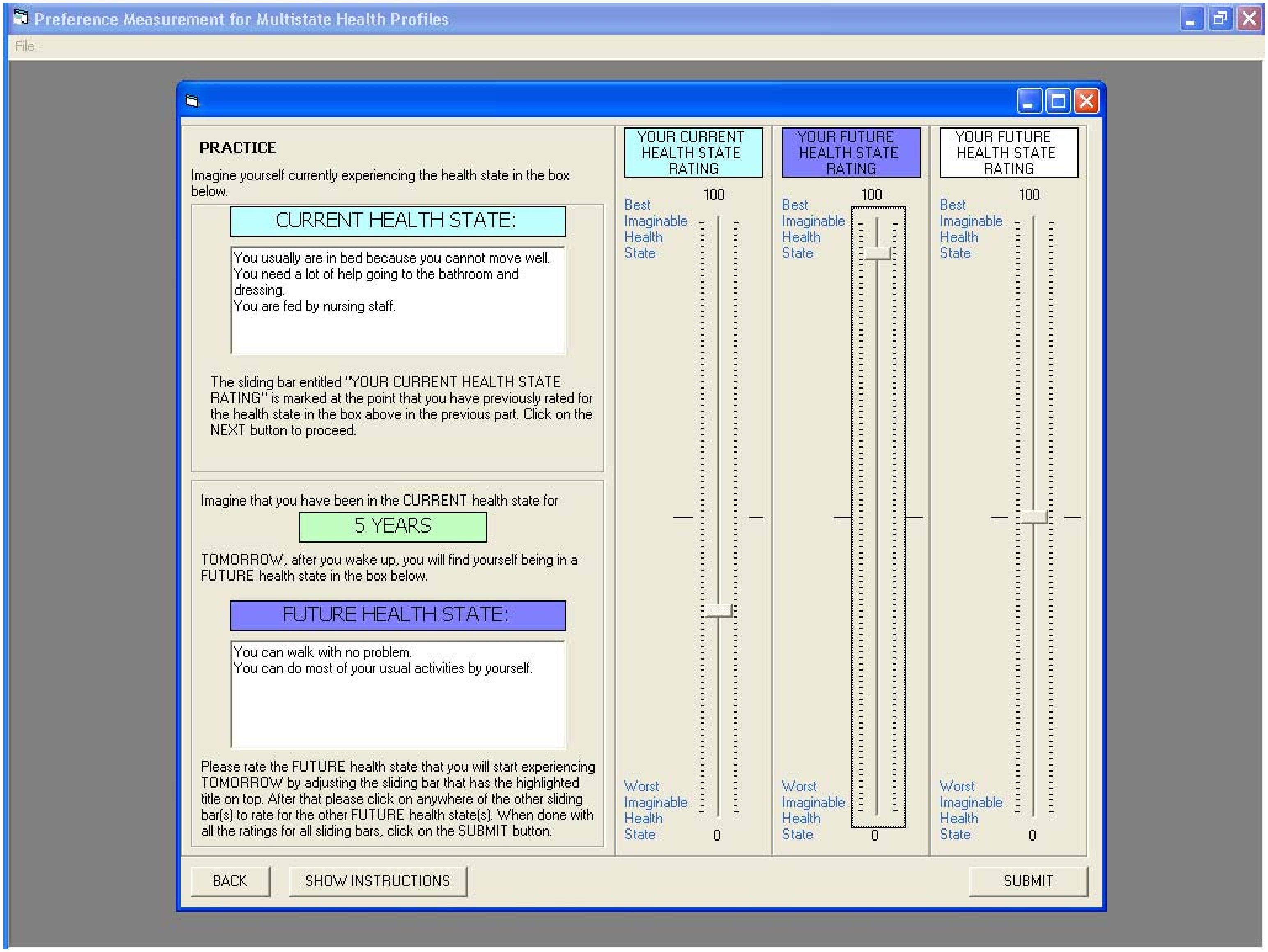

Preference Measurement for Multistate Health Profiles
File
Health States Selection
* For experimenter use only: Input VAS score for each health state. Click on SUBMIT button when finish.
Health State
1
2
3
4
5
6
7
8
9
10
Health State
11
12
13
14
15
16
17
18
19
20
SUBMIT

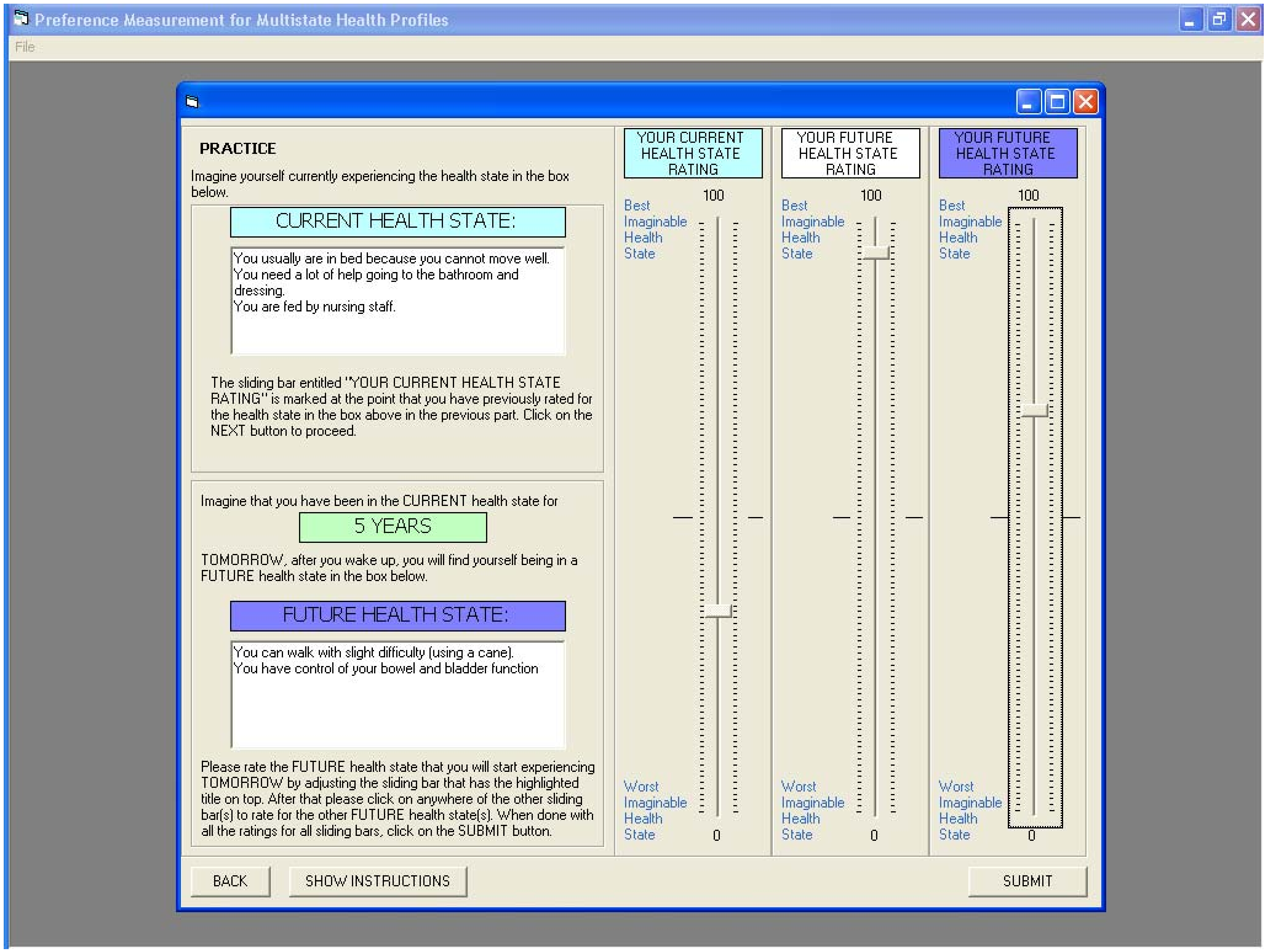

Preference Measurement for Multistate Health Profiles
File
Health States Selection
* For experimenter use only: Review the selected health states and click on SUBMIT button when concur.
Health State
Standardized Score
Reference Score
A 1 100 100
B 5 84 83.83334
C 7 65 67.66666
D 9 52 51.5
E 12 32 35.33333
F 15 19 19.16667
G 19 3 3
see excel table
SUBMIT

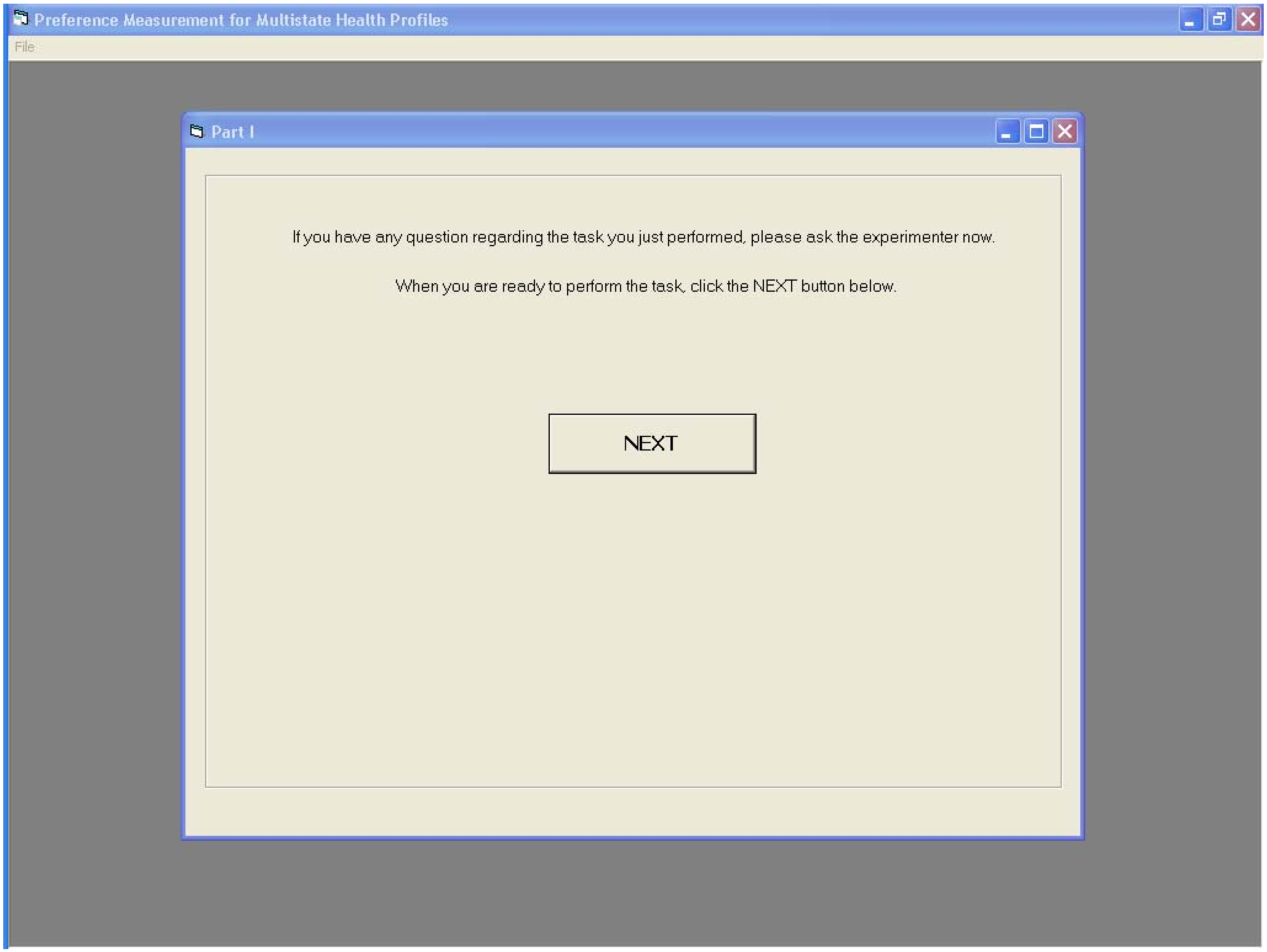

Preference Measurement for Multistate Health Profiles
File
Instructions
PART I
Instructions
In each of the following screens, you will see several hypothetical health scenarios. Each scenario is composed of two health states: your hypothetical CURRENT health state and up to 4 different FUTURE health states. Please read the description of each health scenario carefully.
In addition, the DURATION that you have been experiencing your CURRENT health state is provided. Your FUTURE health state will start on the next day.
First, you will see the hypothetical CURRENT health state along with a slide bar that is already marked at the point that you indicated previously. Remember that the scale ranges from 0 to 100 (given that 0 is the worst imaginable state and 100 is the best imaginable state). Please re-read the description of the CURRENT health state carefully.
Then, you will be presented with one or more hypothetical scenarios, each describing a hypothetical FUTURE health state which you will start experiencing tomorrow immediately after the CURRENT health state which you experienced for a specific period of time presented to you in the scenario. You will be asked to rate how you feel about each hypothetical FUTURE health state by adjusting a new sliding bar corresponding to each FUTURE health state on the right hand side of the screen.
The higher value you put on the scale means that the more preferable the health state is to you.
Remember that in this part, you are asked to rate the FUTURE health state given that you have been in the CURRENT health state for a specific period of time, and NOT the combination of the two health states in the scenario.
Please take your time and adjust the sliding bar(s) until you feel your rating(s) accurately reflects how you feel. When you are done, please click on the SUBMIT button to go to the next set of health scenario.
Click on NEXT button to start.
NEXT

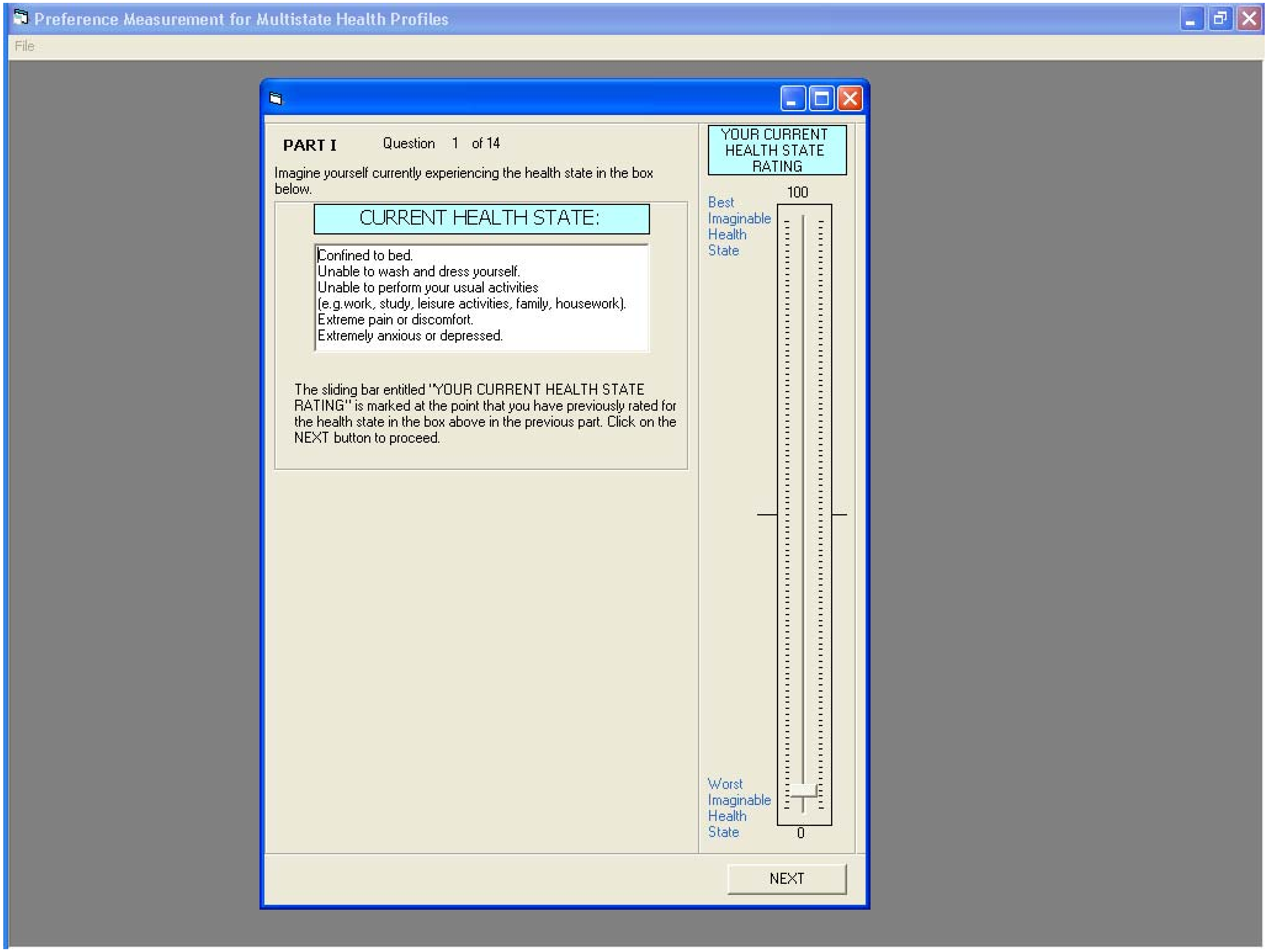

Preference Measurement for Multistate Health Profiles
File
PRACTICE
Imagine yourself currently experiencing the health state in the box below.
CURRENT HEALTH STATE:
You usually are in bed because you cannot move well.
You need a lot of help going to the bathroom and dressing.
You are fed by nursing staff.
The sliding bar entitled "YOUR CURRENT HEALTH STATE RATING" is marked at the point that you have previously rated for the health state in the box above in the previous part. Click on the NEXT button to proceed.
YOUR CURRENT HEALTH STATE RATING
100
Best Imaginable Health State
Worst Imaginable Health State
0
BACK
NEXT
ios. Each scenario is composed of
FUTURE health states. Please read
th state is provided. Your FUTURE
that is already marked at the point
iven that 0 is the worst imaginable
e CURRENT health state carefully.
ribing a hypothetical FUTURE
URRENT health state which you
will be asked to rate how you feel
responding to each FUTURE health
th state is to you.
n that you have been in the
the two health states in the
accurately reflects how you feel.
health scenario.
NEXT

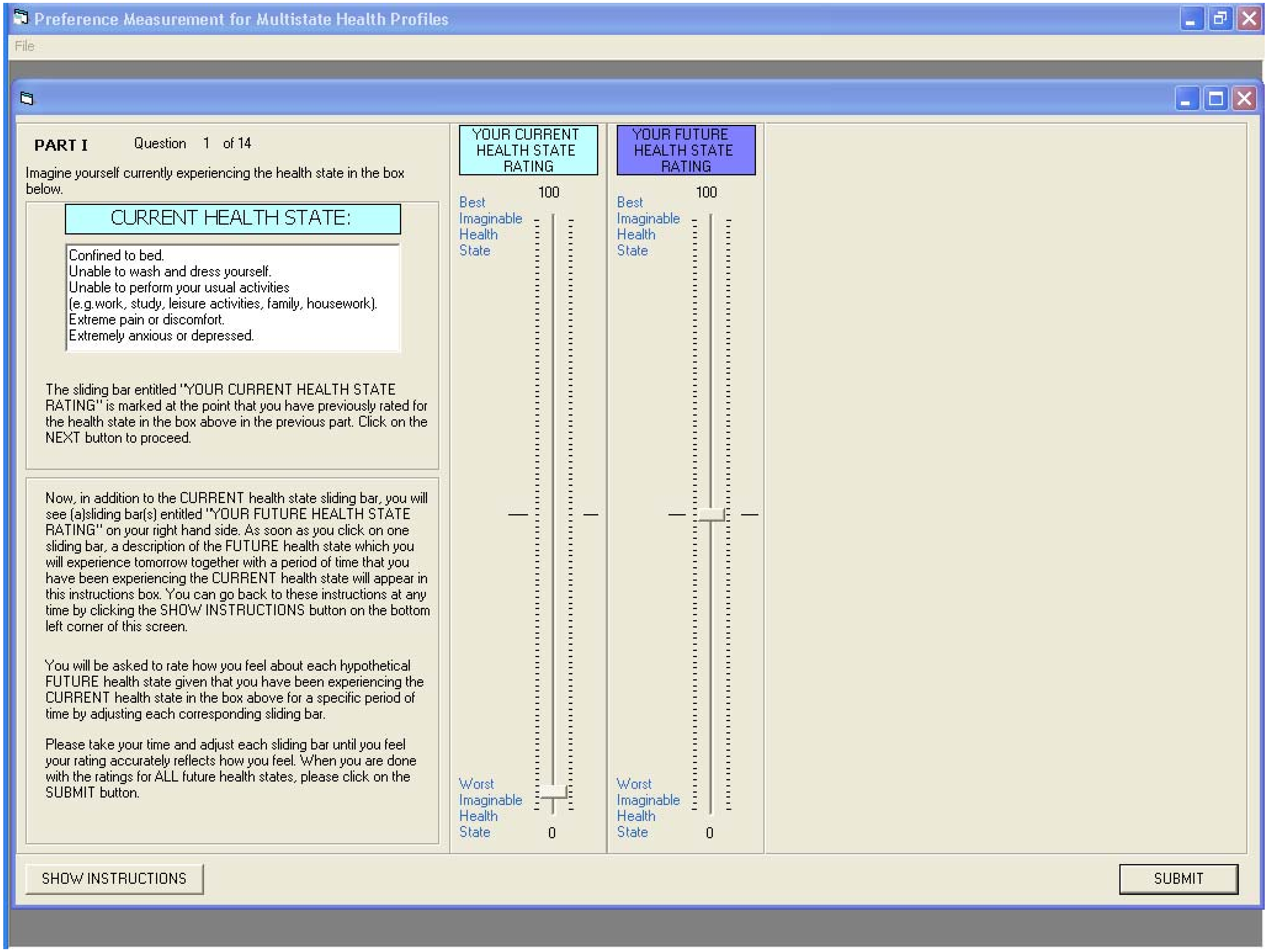

Preference Measurement for Multistate Health Profiles
File
PRACTICE
Imagine yourself currently experiencing the health state in the box below.
CURRENT HEALTH STATE:
You usually are in bed because you cannot move well.
You need a lot of help going to the bathroom and dressing.
You are fed by nursing staff.
The sliding bar entitled "YOUR CURRENT HEALTH STATE RATING" is marked at the point that you have previously rated for the health state in the box above in the previous part. Click on the NEXT button to proceed.
Now, in addition to the CURRENT health state sliding bar, you will see (a)sliding bar(s) entitled "YOUR FUTURE HEALTH STATE RATING" on your right hand side. As soon as you click on one sliding bar, a description of the FUTURE health state which you will experience tomorrow together with a period of time that you have been experiencing the CURRENT health state will appear in this instructions box. You can go back to these instructions at any time by clicking the SHOW INSTRUCTIONS button on the bottom left corner of this screen.
You will be asked to rate how you feel about each hypothetical FUTURE health state given that you have been experiencing the CURRENT health state in the box above for a specific period of time by adjusting each corresponding sliding bar.
Please take your time and adjust each sliding bar until you feel your rating accurately reflects how you feel. When you are done with the ratings for ALL future health states, please click on the SUBMIT button.
YOUR CURRENT HEALTH STATE RATING
YOUR FUTURE HEALTH STATE RATING
YOUR FUTURE HEALTH STATE RATING
100
Best Imaginable Health State
Worst Imaginable Health State
0
BACK
SHOW INSTRUCTIONS
SUBMIT


Preference Measurement for Multistate Health Profiles
File
PRACTICE
Imagine yourself currently experiencing the health state in the box below.
CURRENT HEALTH STATE:
You usually are in bed because you cannot move well.
You need a lot of help going to the bathroom and dressing.
You are fed by nursing staff.
The sliding bar entitled "YOUR CURRENT HEALTH STATE RATING" is marked at the point that you have previously rated for the health state in the box above in the previous part. Click on the NEXT button to proceed.
Imagine that you have been in the CURRENT health state for
5 YEARS
TOMORROW, after you wake up, you will find yourself being in a FUTURE health state in the box below.
FUTURE HEALTH STATE:
You can walk with no problem.
You can do most of your usual activities by yourself.
Please rate the FUTURE health state that you will start experiencing TOMORROW by adjusting the sliding bar that has the highlighted title on top. After that please click on anywhere of the other sliding bar(s) to rate for the other FUTURE health state(s). When done with all the ratings for all sliding bars, click on the SUBMIT button.
YOUR CURRENT HEALTH STATE RATING
YOUR FUTURE HEALTH STATE RATING
YOUR FUTURE HEALTH STATE RATING
100
Best Imaginable Health State
Worst Imaginable Health State
0
BACK
SHOW INSTRUCTIONS
SUBMIT

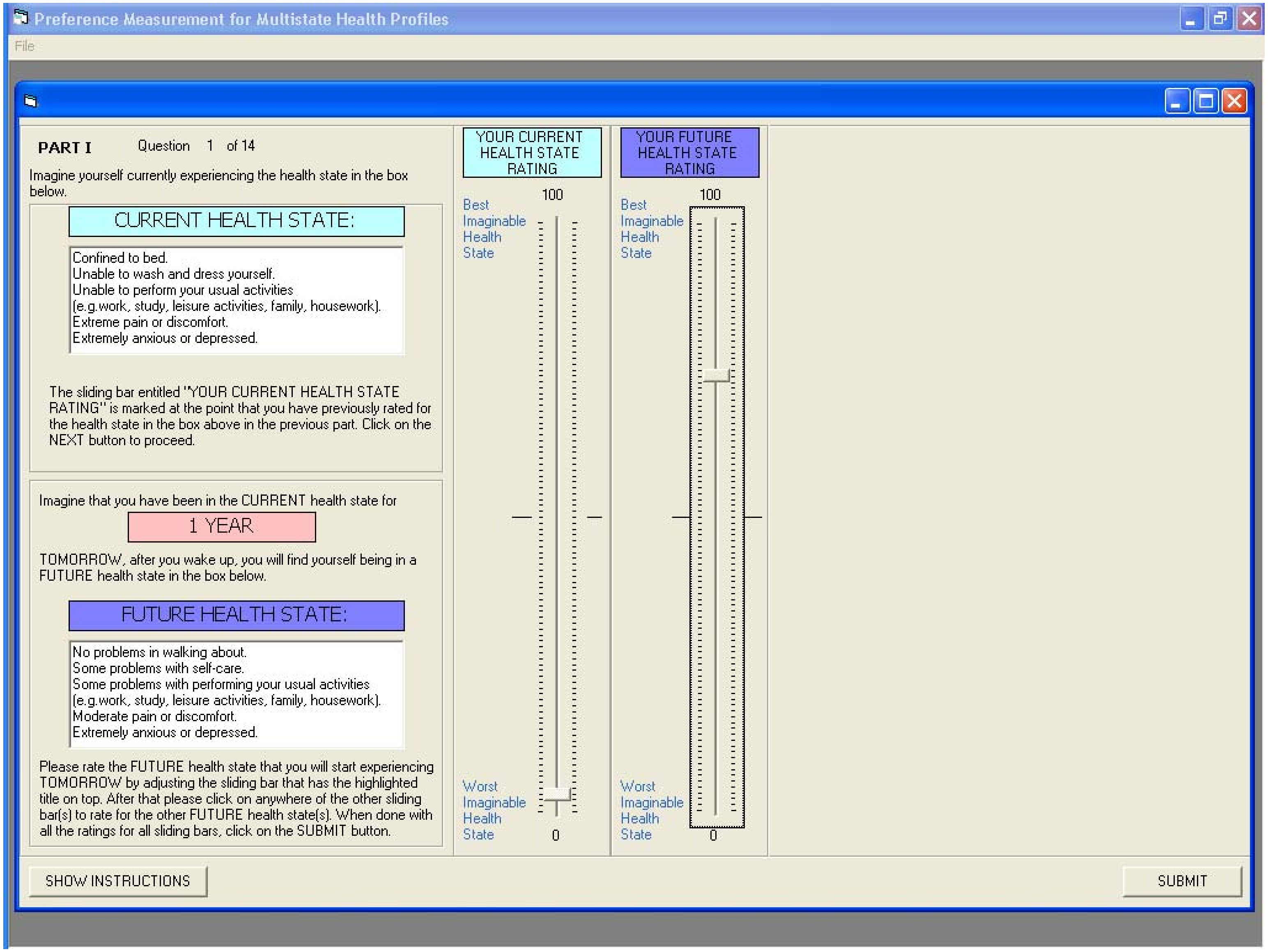

Preference Measurement for Multistate Health Profiles
File
PRACTICE
Imagine yourself currently experiencing the health state in the box below.
CURRENT HEALTH STATE:
You usually are in bed because you cannot move well.
You need a lot of help going to the bathroom and dressing.
You are fed by nursing staff.
The sliding bar entitled "YOUR CURRENT HEALTH STATE RATING" is marked at the point that you have previously rated for the health state in the box above in the previous part. Click on the NEXT button to proceed.
Imagine that you have been in the CURRENT health state for
5 YEARS
TOMORROW, after you wake up, you will find yourself being in a FUTURE health state in the box below.
FUTURE HEALTH STATE:
You can walk with slight difficulty (using a cane).
You have control of your bowel and bladder function
Please rate the FUTURE health state that you will start experiencing TOMORROW by adjusting the sliding bar that has the highlighted title on top. After that please click on anywhere of the other sliding bar(s) to rate for the other FUTURE health state(s). When done with all the ratings for all sliding bars, click on the SUBMIT button.
YOUR CURRENT HEALTH STATE RATING
YOUR FUTURE HEALTH STATE RATING
YOUR FUTURE HEALTH STATE RATING
100
Best Imaginable Health State
Worst Imaginable Health State
0
BACK
SHOW INSTRUCTIONS
SUBMIT

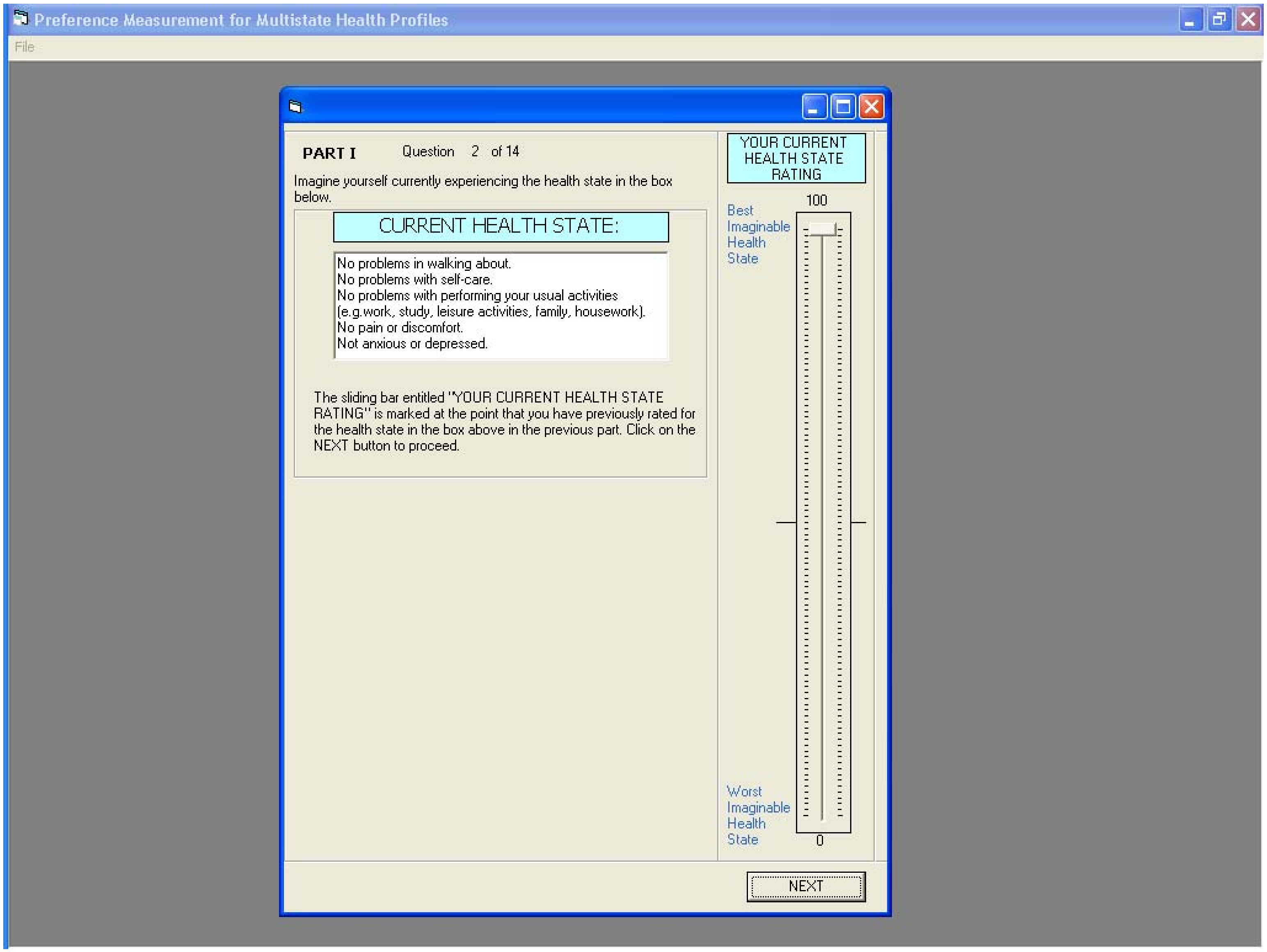

Preference Measurement for Multistate Health Profiles
File
Part I
If you have any question regarding the task you just performed, please ask the experimenter now.
When you are ready to perform the task, click the NEXT button below.
NEXT

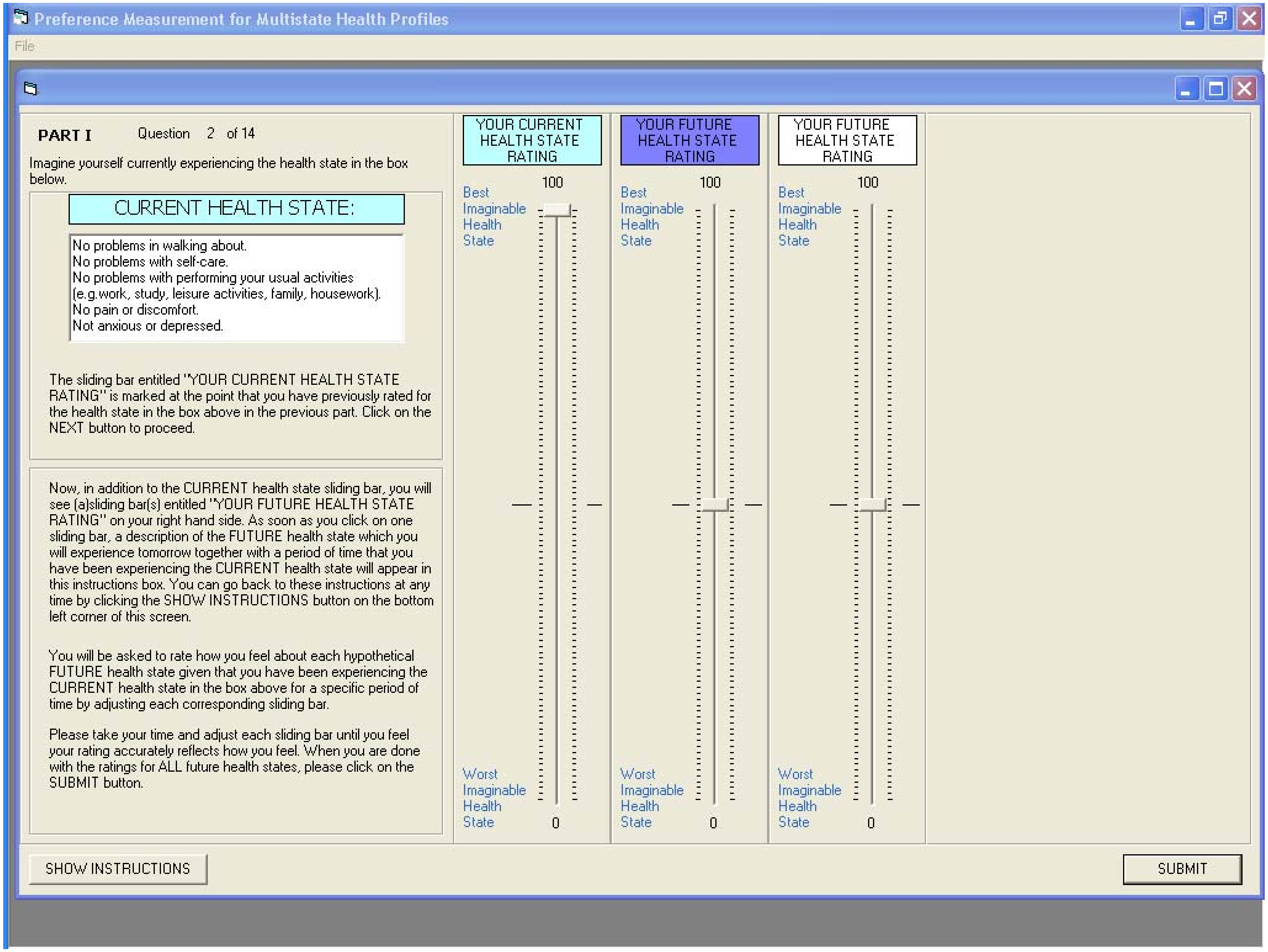

Preference Measurement for Multistate Health Profiles
File
PART I
Question 1 of 14
Imagine yourself currently experiencing the health state in the box below.
CURRENT HEALTH STATE:
Confined to bed.
Unable to wash and dress yourself.
Unable to perform your usual activities
(e.g.work, study, leisure activities, family, housework).
Extreme pain or discomfort.
Extremely anxious or depressed.
The sliding bar entitled "YOUR CURRENT HEALTH STATE RATING" is marked at the point that you have previously rated for the health state in the box above in the previous part. Click on the NEXT button to proceed.
YOUR CURRENT HEALTH STATE RATING
100
Best Imaginable Health State
Worst Imaginable Health State
0
NEXT

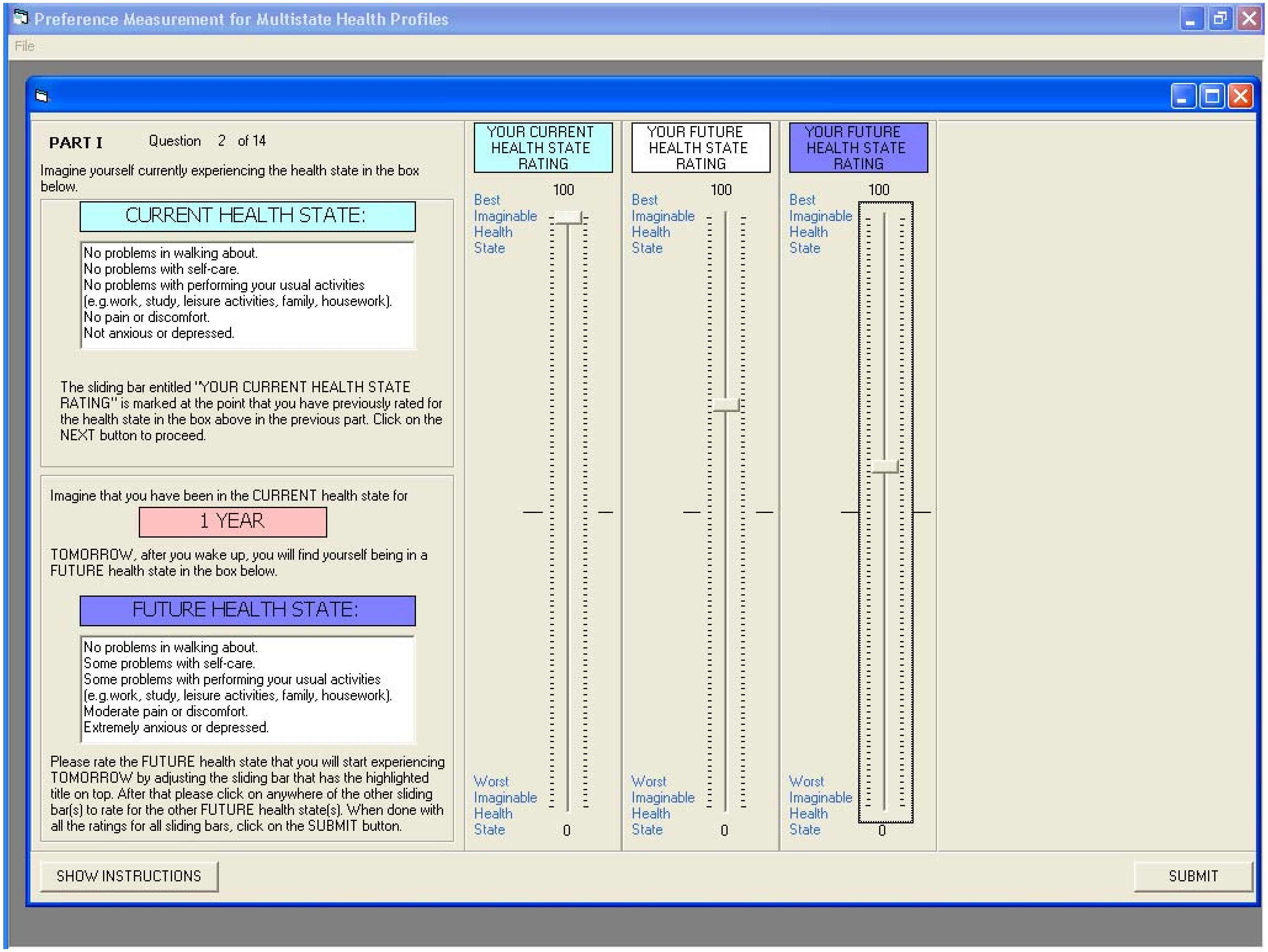


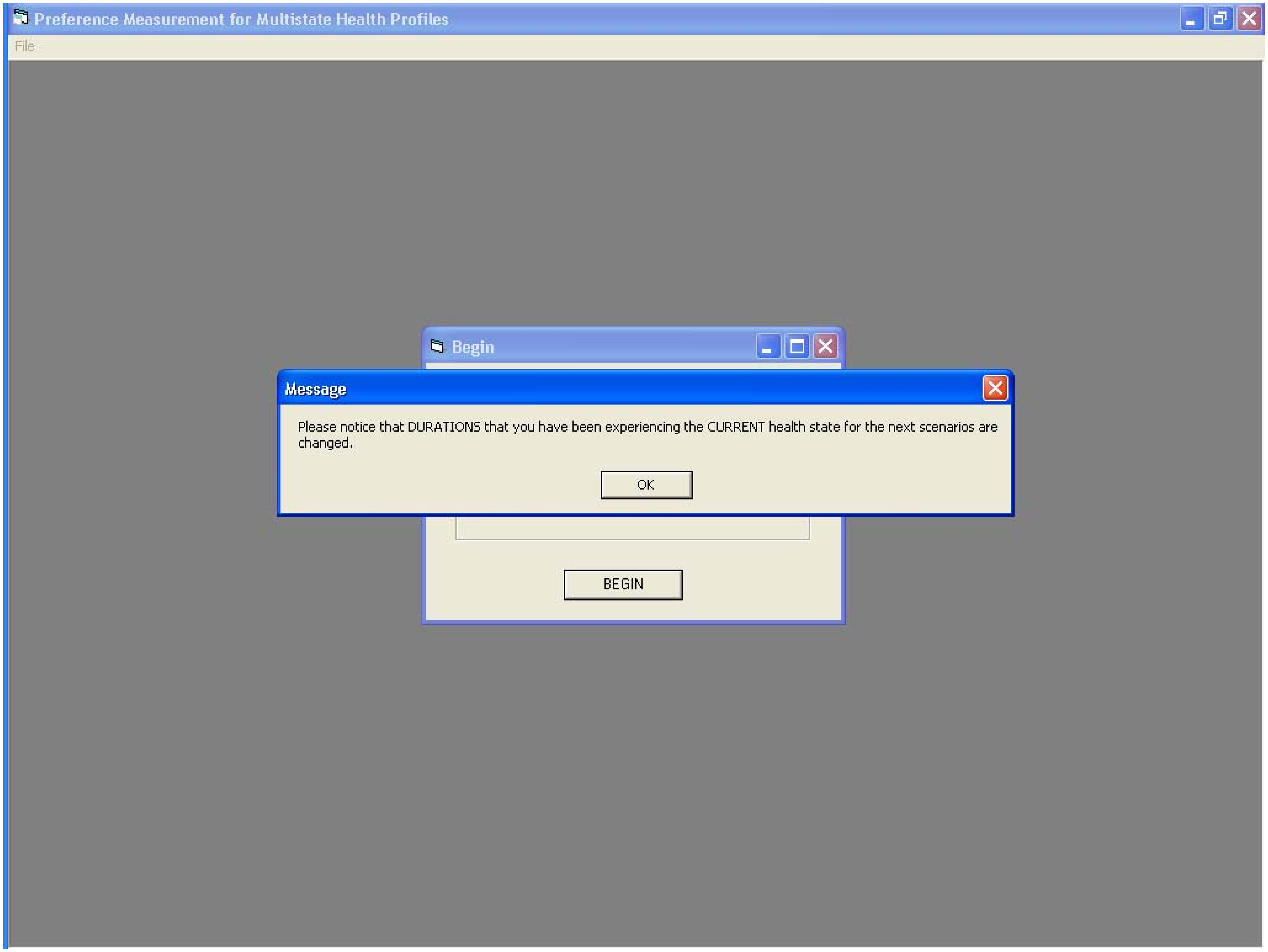
Preference Measurement for Multistate Health Profiles
File
PART I
Question 1 of 14
Imagine yourself currently experiencing the health state in the box below.
CURRENT HEALTH STATE:
Confined to bed.
Unable to wash and dress yourself.
Unable to perform your usual activities
(e.g.work, study, leisure activities, family, housework).
Extreme pain or discomfort.
Extremely anxious or depressed.
The sliding bar entitled "YOUR CURRENT HEALTH STATE RATING" is marked at the point that you have previously rated for the health state in the box above in the previous part. Click on the NEXT button to proceed.
Now, in addition to the CURRENT health state sliding bar, you will see (a)sliding bar(s) entitled "YOUR FUTURE HEALTH STATE RATING" on your right hand side. As soon as you click on one sliding bar, a description of the FUTURE health state which you will experience tomorrow together with a period of time that you have been experiencing the CURRENT health state will appear in this instructions box. You can go back to these instructions at any time by clicking the SHOW INSTRUCTIONS button on the bottom left corner of this screen.
You will be asked to rate how you feel about each hypothetical FUTURE health state given that you have been experiencing the CURRENT health state in the box above for a specific period of time by adjusting each corresponding sliding bar.
Please take your time and adjust each sliding bar until you feel your rating accurately reflects how you feel. When you are done with the ratings for ALL future health states, please click on the SUBMIT button.
YOUR CURRENT HEALTH STATE RATING
100
Best Imaginable Health State
Worst Imaginable Health State
0
YOUR FUTURE HEALTH STATE RATING
100
Best Imaginable Health State
Worst Imaginable Health State
0
SHOW INSTRUCTIONS
SUBMIT

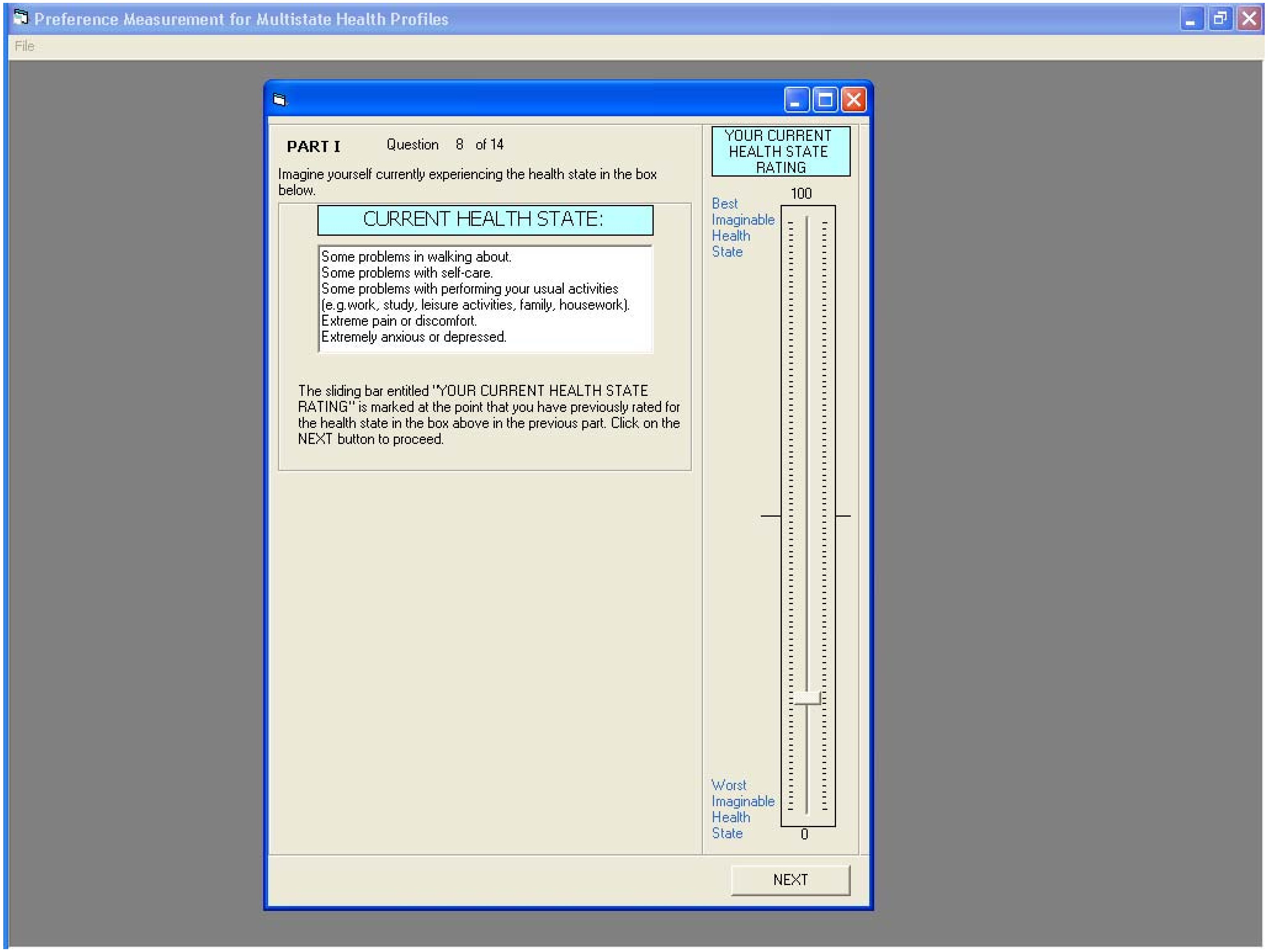

Preference Measurement for Multistate Health Profiles
File
PART I
Question 1 of 14
Imagine yourself currently experiencing the health state in the box below.
CURRENT HEALTH STATE:
Confined to bed.
Unable to wash and dress yourself.
Unable to perform your usual activities
(e.g.work, study, leisure activities, family, housework).
Extreme pain or discomfort.
Extremely anxious or depressed.
The sliding bar entitled "YOUR CURRENT HEALTH STATE RATING" is marked at the point that you have previously rated for the health state in the box above in the previous part. Click on the NEXT button to proceed.
Imagine that you have been in the CURRENT health state for
1 YEAR
TOMORROW, after you wake up, you will find yourself being in a FUTURE health state in the box below.
FUTURE HEALTH STATE:
No problems in walking about.
Some problems with self-care.
Some problems with performing your usual activities
(e.g.work, study, leisure activities, family, housework).
Moderate pain or discomfort.
Extremely anxious or depressed.
Please rate the FUTURE health state that you will start experiencing TOMORROW by adjusting the sliding bar that has the highlighted title on top. After that please click on anywhere of the other sliding bar(s) to rate for the other FUTURE health state(s). When done with all the ratings for all sliding bars, click on the SUBMIT button.
YOUR CURRENT HEALTH STATE RATING
100
Best Imaginable Health State
Worst Imaginable Health State
0
YOUR FUTURE HEALTH STATE RATING
100
Best Imaginable Health State
Worst Imaginable Health State
0
SHOW INSTRUCTIONS
SUBMIT

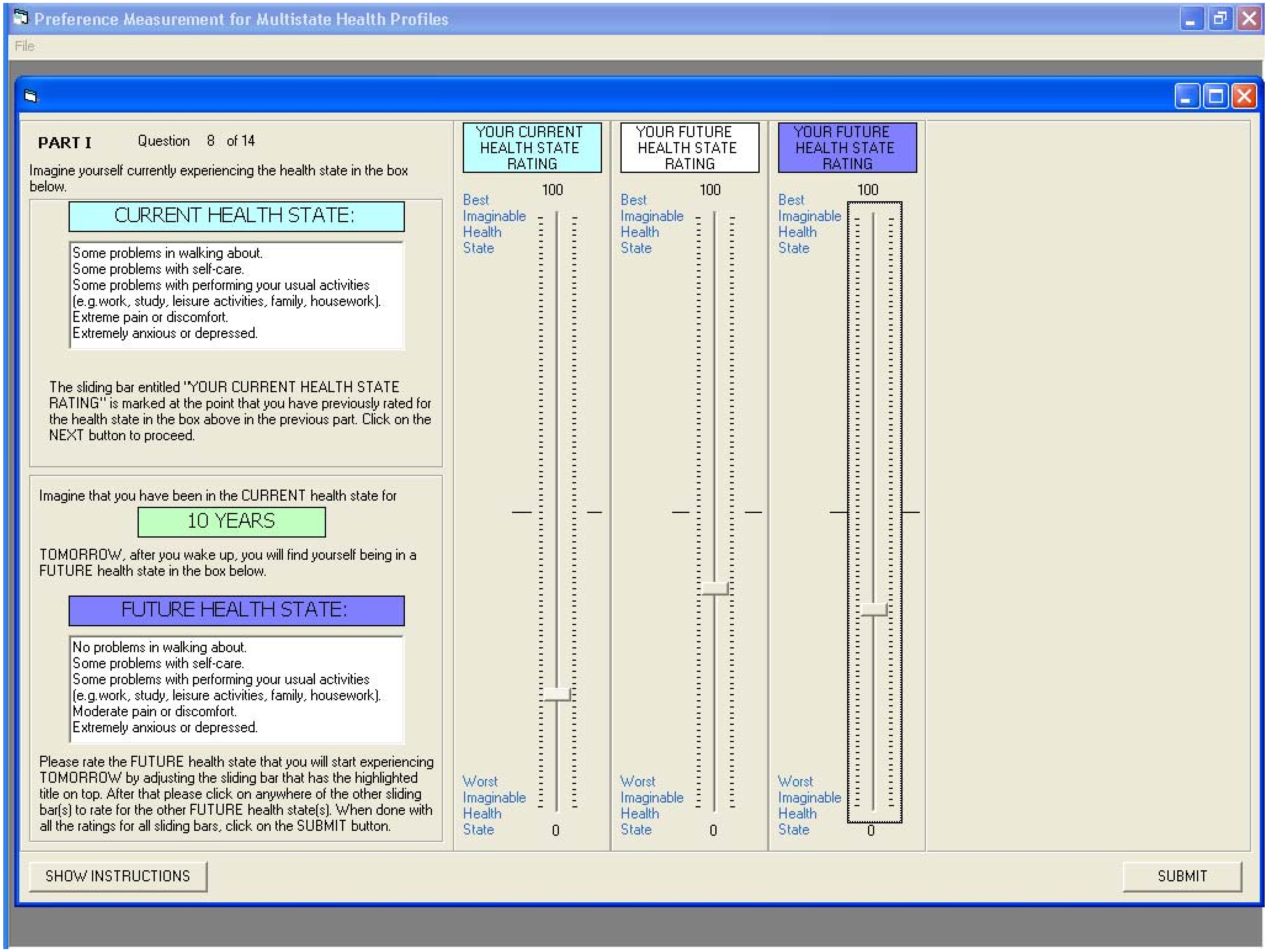

Preference Measurement for Multistate Health Profiles
File
PART I
Question 2 of 14
Imagine yourself currently experiencing the health state in the box below.
CURRENT HEALTH STATE:
No problems in walking about.
No problems with self-care.
No problems with performing your usual activities (e.g.work, study, leisure activities, family, housework).
No pain or discomfort.
Not anxious or depressed.
The sliding bar entitled "YOUR CURRENT HEALTH STATE RATING" is marked at the point that you have previously rated for the health state in the box above in the previous part. Click on the NEXT button to proceed.
YOUR CURRENT HEALTH STATE RATING
100
Best Imaginable Health State
Worst Imaginable Health State
0
NEXT

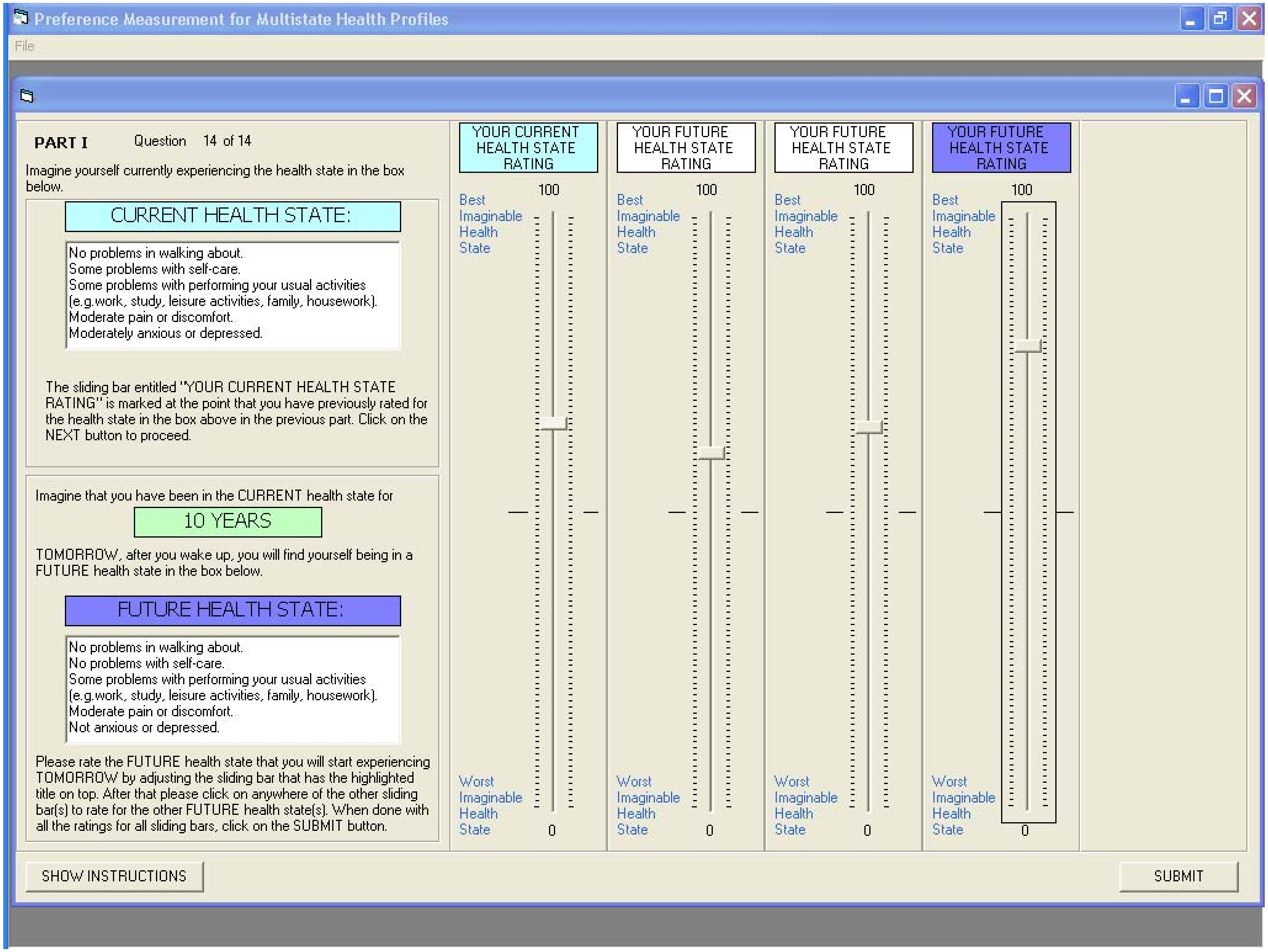

Preference Measurement for Multistate Health Profiles
File
PART I
Question 2 of 14
Imagine yourself currently experiencing the health state in the box below.
CURRENT HEALTH STATE:
No problems in walking about.
No problems with self-care.
No problems with performing your usual activities (e.g.work, study, leisure activities, family, housework).
No pain or discomfort.
Not anxious or depressed.
The sliding bar entitled "YOUR CURRENT HEALTH STATE RATING" is marked at the point that you have previously rated for the health state in the box above in the previous part. Click on the NEXT button to proceed.
Now, in addition to the CURRENT health state sliding bar, you will see (a)sliding bar(s) entitled "YOUR FUTURE HEALTH STATE RATING" on your right hand side. As soon as you click on one sliding bar, a description of the FUTURE health state which you will experience tomorrow together with a period of time that you have been experiencing the CURRENT health state will appear in this instructions box. You can go back to these instructions at any time by clicking the SHOW INSTRUCTIONS button on the bottom left corner of this screen.
You will be asked to rate how you feel about each hypothetical FUTURE health state given that you have been experiencing the CURRENT health state in the box above for a specific period of time by adjusting each corresponding sliding bar.
Please take your time and adjust each sliding bar until you feel your rating accurately reflects how you feel. When you are done with the ratings for ALL future health states, please click on the SUBMIT button.
YOUR CURRENT HEALTH STATE RATING
YOUR FUTURE HEALTH STATE RATING
YOUR FUTURE HEALTH STATE RATING
Best Imaginable Health State
100
Worst Imaginable Health State
0
SHOW INSTRUCTIONS
SUBMIT

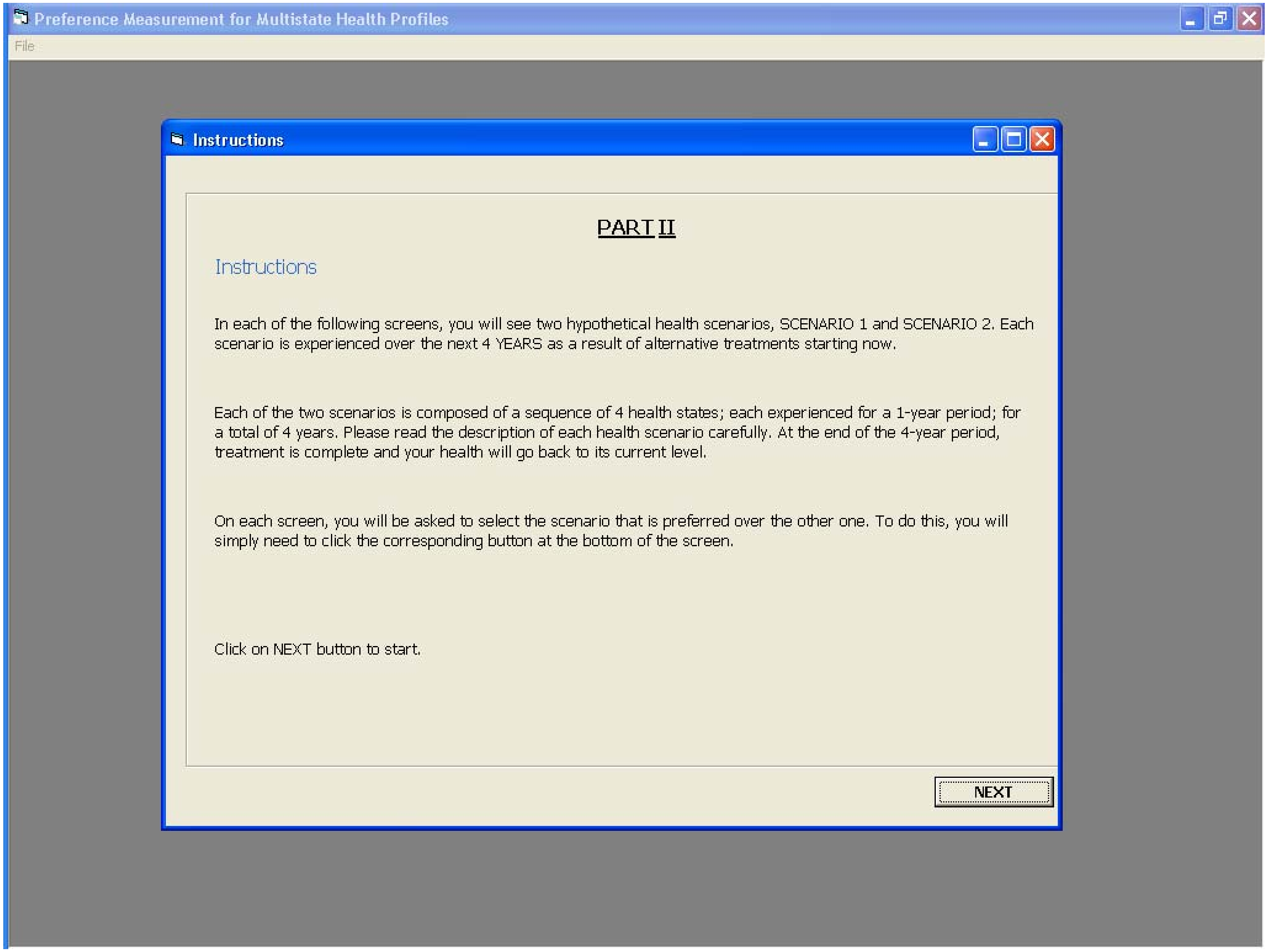

Preference Measurement for Multistate Health Profiles
File
PART I
Question 2 of 14
Imagine yourself currently experiencing the health state in the box below.
CURRENT HEALTH STATE:
No problems in walking about.
No problems with self-care.
No problems with performing your usual activities (e.g.work, study, leisure activities, family, housework).
No pain or discomfort.
Not anxious or depressed.
The sliding bar entitled "YOUR CURRENT HEALTH STATE RATING" is marked at the point that you have previously rated for the health state in the box above in the previous part. Click on the NEXT button to proceed.
Imagine that you have been in the CURRENT health state for
1 YEAR
TOMORROW, after you wake up, you will find yourself being in a FUTURE health state in the box below.
FUTURE HEALTH STATE:
No problems in walking about.
Some problems with self-care.
Some problems with performing your usual activities (e.g.work, study, leisure activities, family, housework).
Moderate pain or discomfort.
Extremely anxious or depressed.
Please rate the FUTURE health state that you will start experiencing TOMORROW by adjusting the sliding bar that has the highlighted title on top. After that please click on anywhere of the other sliding bar(s) to rate for the other FUTURE health state(s). When done with all the ratings for all sliding bars, click on the SUBMIT button.
YOUR CURRENT HEALTH STATE RATING
YOUR FUTURE HEALTH STATE RATING
YOUR FUTURE HEALTH STATE RATING
100
Best Imaginable Health State
Worst Imaginable Health State
0
SHOW INSTRUCTIONS
SUBMIT

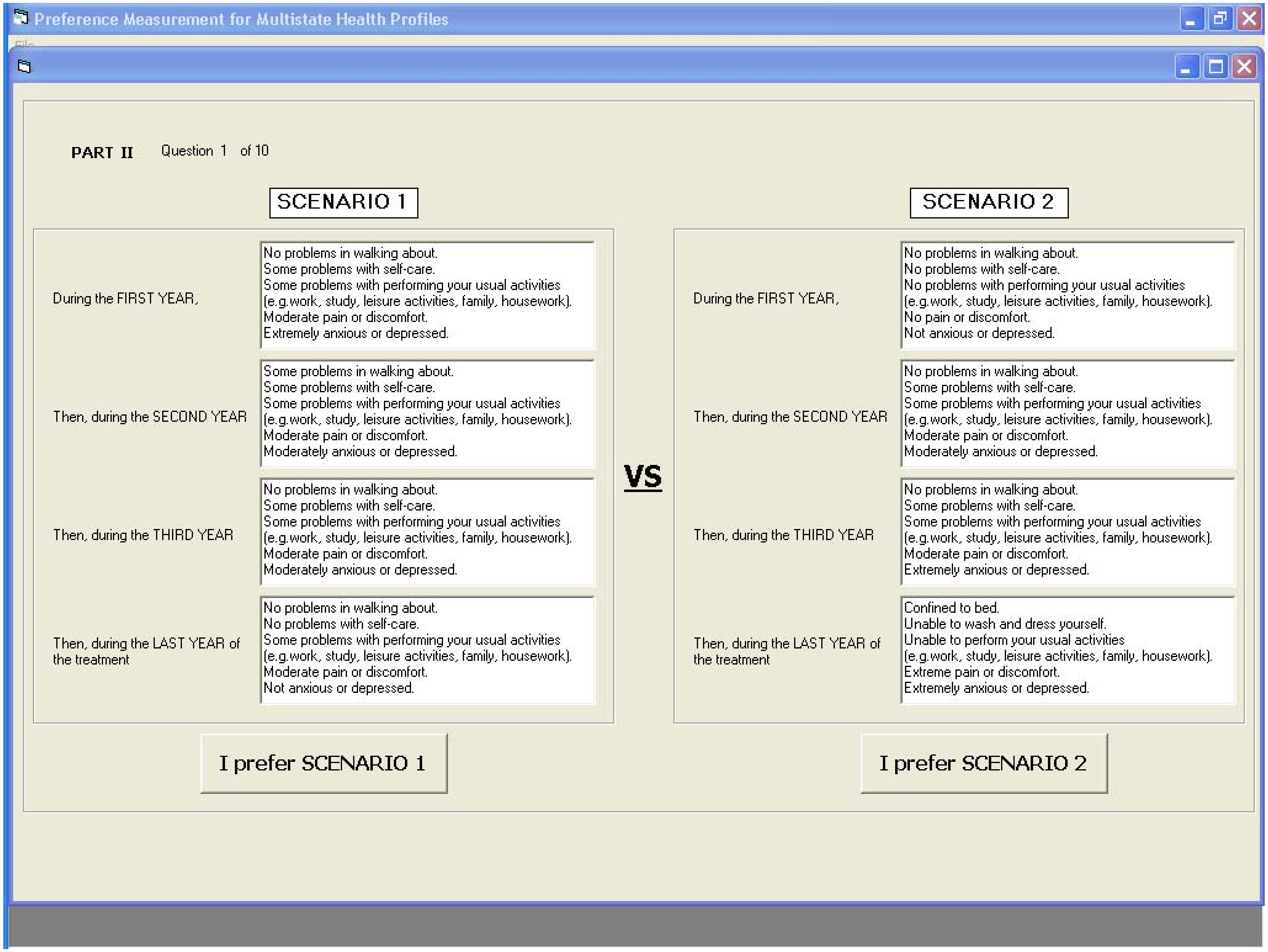

Preference Measurement for Multistate Health Profiles
File
Begin
Message
Please notice that DURATIONS that you have been experiencing the CURRENT health state for the next scenarios are changed.
OK
BEGIN


Preference Measurement for Multistate Health Profiles
File
PART I
Question 8 of 14
Imagine yourself currently experiencing the health state in the box below.
CURRENT HEALTH STATE:
Some problems in walking about.
Some problems with self-care.
Some problems with performing your usual activities (e.g.work, study, leisure activities, family, housework).
Extreme pain or discomfort.
Extremely anxious or depressed.
The sliding bar entitled "YOUR CURRENT HEALTH STATE RATING" is marked at the point that you have previously rated for the health state in the box above in the previous part. Click on the NEXT button to proceed.
YOUR CURRENT HEALTH STATE RATING
100
Best Imaginable Health State
Worst Imaginable Health State
0
NEXT


Preference Measurement for Multistate Health Profiles
File
PART I
Question 8 of 14
Imagine yourself currently experiencing the health state in the box below.
CURRENT HEALTH STATE:
Some problems in walking about.
Some problems with self-care.
Some problems with performing your usual activities (e.g.work, study, leisure activities, family, housework).
Extreme pain or discomfort.
Extremely anxious or depressed.
The sliding bar entitled "YOUR CURRENT HEALTH STATE RATING" is marked at the point that you have previously rated for the health state in the box above in the previous part. Click on the NEXT button to proceed.
Imagine that you have been in the CURRENT health state for
10 YEARS
TOMORROW, after you wake up, you will find yourself being in a FUTURE health state in the box below.
FUTURE HEALTH STATE:
No problems in walking about.
Some problems with self-care.
Some problems with performing your usual activities (e.g.work, study, leisure activities, family, housework).
Moderate pain or discomfort.
Extremely anxious or depressed.
Please rate the FUTURE health state that you will start experiencing TOMORROW by adjusting the sliding bar that has the highlighted title on top. After that please click on anywhere of the other sliding bar(s) to rate for the other FUTURE health state(s). When done with all the ratings for all sliding bars, click on the SUBMIT button.
YOUR CURRENT HEALTH STATE RATING
YOUR FUTURE HEALTH STATE RATING
YOUR FUTURE HEALTH STATE RATING
100
Best Imaginable Health State
Worst Imaginable Health State
0
SHOW INSTRUCTIONS
SUBMIT


Preference Measurement for Multistate Health Profiles
File
PART I
Question 14 of 14
Imagine yourself currently experiencing the health state in the box below.
CURRENT HEALTH STATE:
No problems in walking about.
Some problems with self-care.
Some problems with performing your usual activities (e.g.work, study, leisure activities, family, housework).
Moderate pain or discomfort.
Moderately anxious or depressed.
The sliding bar entitled "YOUR CURRENT HEALTH STATE RATING" is marked at the point that you have previously rated for the health state in the box above in the previous part. Click on the NEXT button to proceed.
Imagine that you have been in the CURRENT health state for
10 YEARS
TOMORROW, after you wake up, you will find yourself being in a FUTURE health state in the box below.
FUTURE HEALTH STATE:
No problems in walking about.
No problems with self-care.
Some problems with performing your usual activities (e.g.work, study, leisure activities, family, housework).
Moderate pain or discomfort.
Not anxious or depressed.
Please rate the FUTURE health state that you will start experiencing TOMORROW by adjusting the sliding bar that has the highlighted title on top. After that please click on anywhere of the other sliding bar(s) to rate for the other FUTURE health state(s). When done with all the ratings for all sliding bars, click on the SUBMIT button.
YOUR CURRENT HEALTH STATE RATING
YOUR FUTURE HEALTH STATE RATING
YOUR FUTURE HEALTH STATE RATING
YOUR FUTURE HEALTH STATE RATING
100
Best Imaginable Health State
Worst Imaginable Health State
0
SHOW INSTRUCTIONS
SUBMIT


Preference Measurement for Multistate Health Profiles
File
Instructions
PART II
Instructions
In each of the following screens, you will see two hypothetical health scenarios, SCENARIO 1 and SCENARIO 2. Each scenario is experienced over the next 4 YEARS as a result of alternative treatments starting now.
Each of the two scenarios is composed of a sequence of 4 health states; each experienced for a 1-year period; for a total of 4 years. Please read the description of each health scenario carefully. At the end of the 4-year period, treatment is complete and your health will go back to its current level.
On each screen, you will be asked to select the scenario that is preferred over the other one. To do this, you will simply need to click the corresponding button at the bottom of the screen.
Click on NEXT button to start.
NEXT

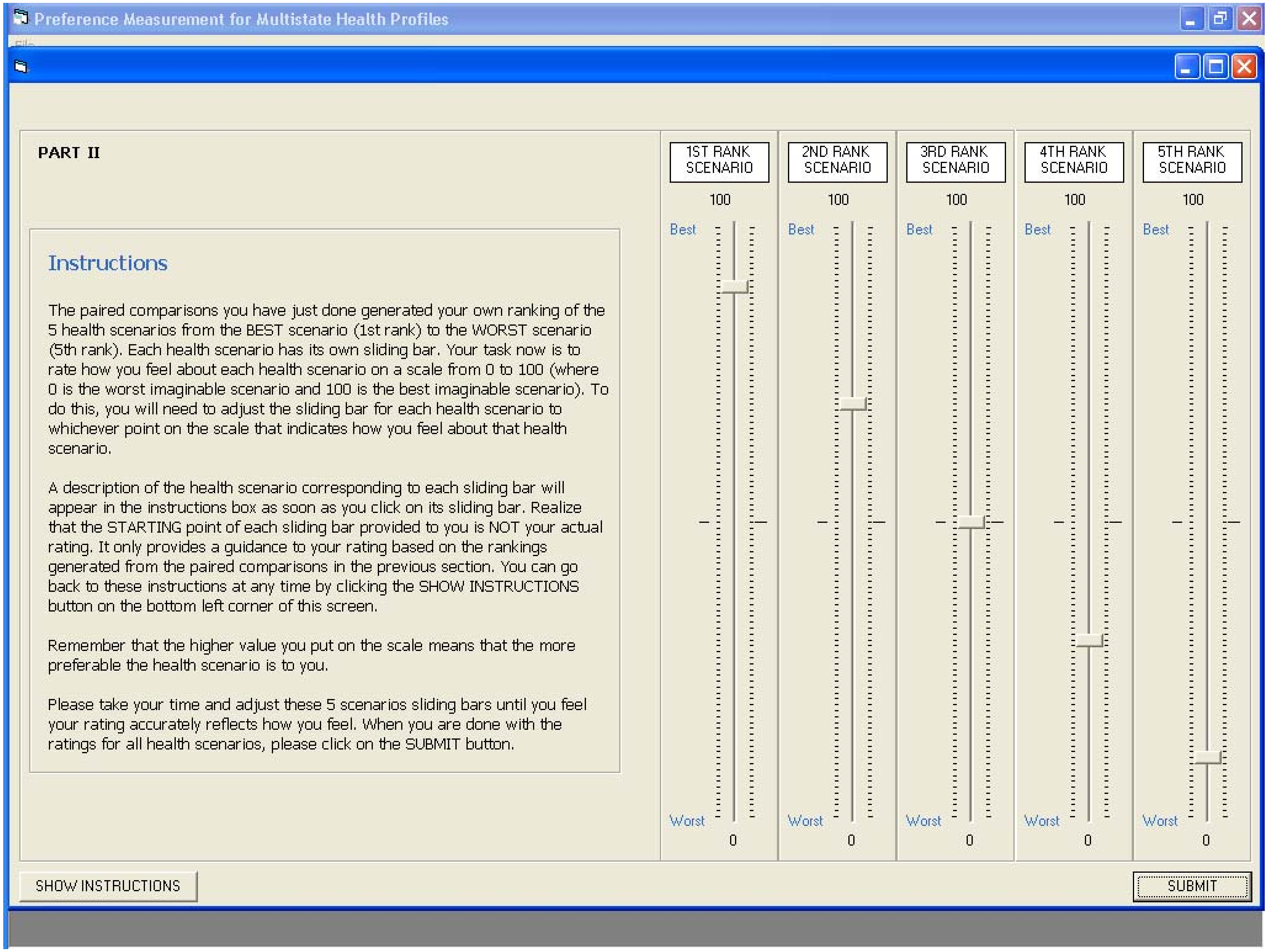


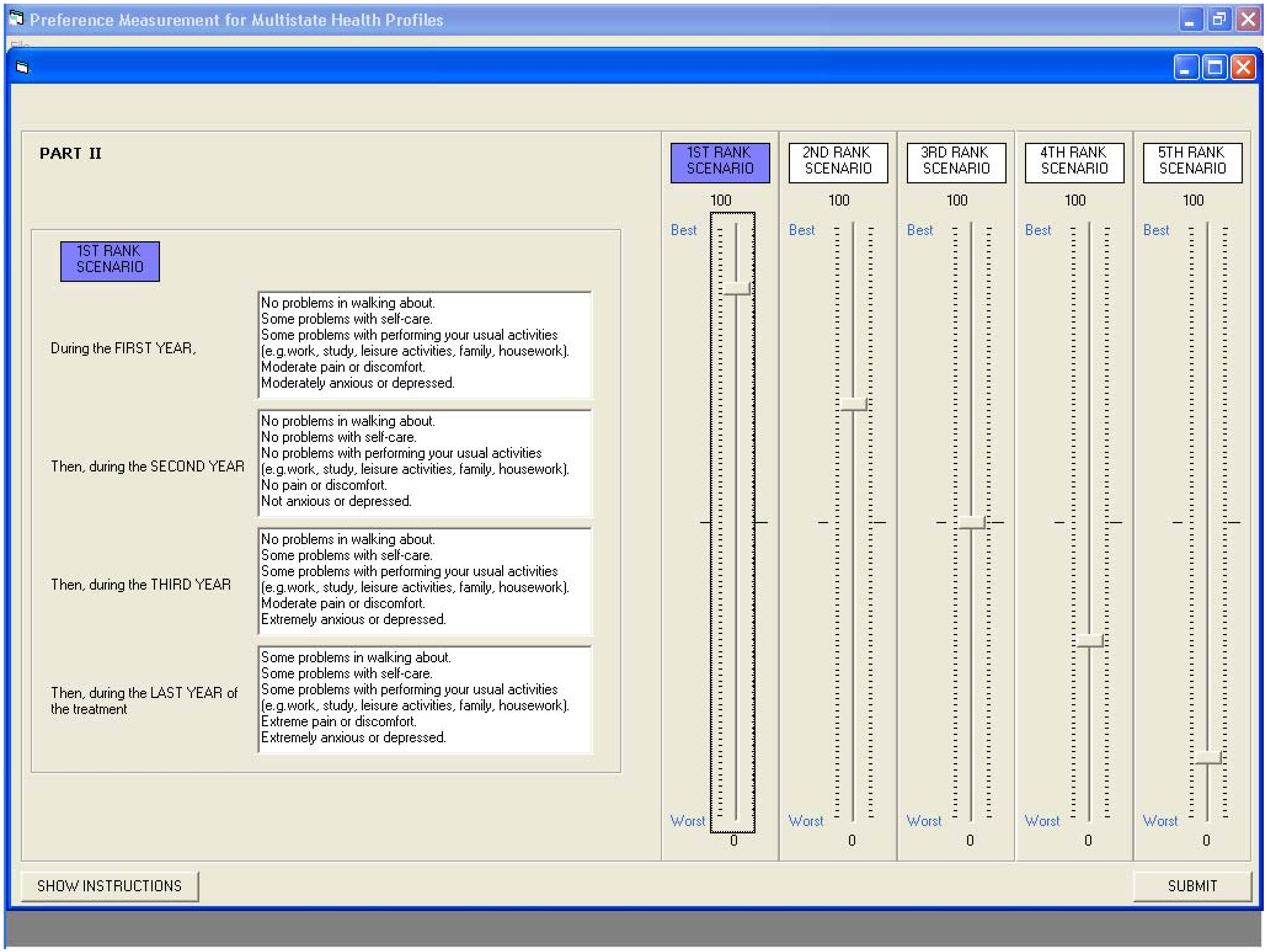
Preference Measurement for Multistate Health Profiles

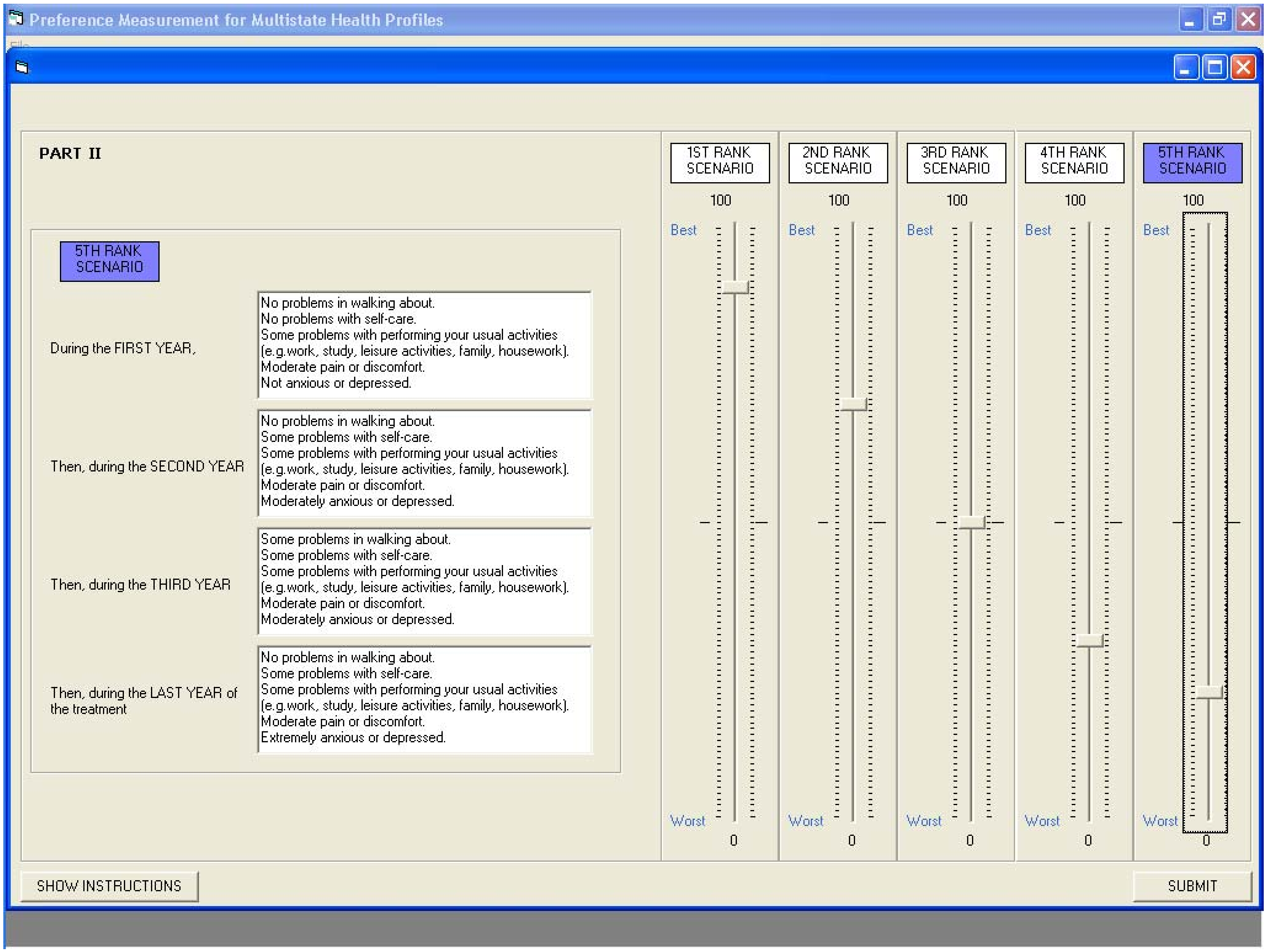
PART II Question 1 of 10

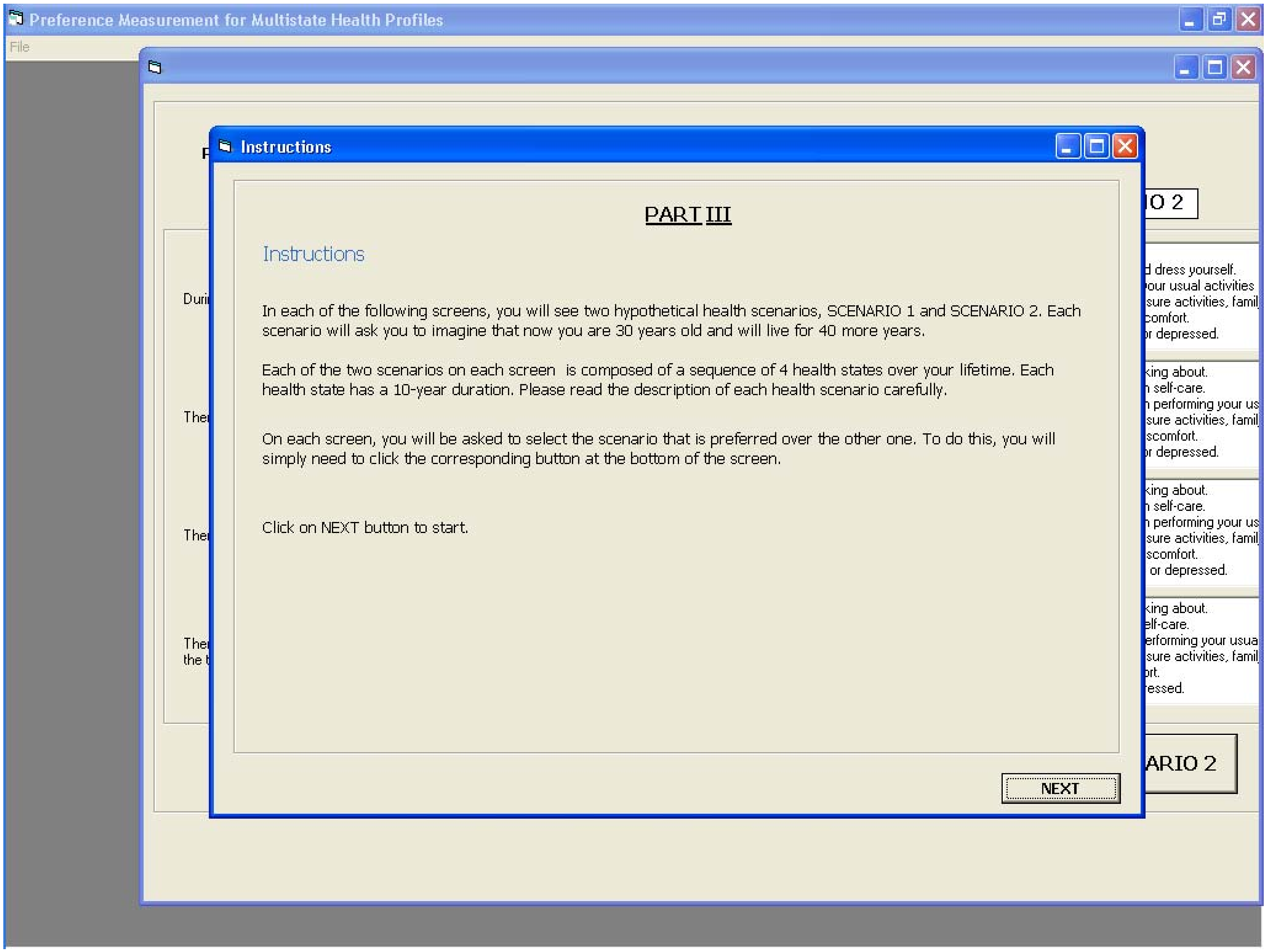
SCENARIO 1		VS	SCENARIO 2	
During the FIRST YEAR,	No problems in walking about. Some problems with self-care. Some problems with performing your usual activities (e.g.work, study, leisure activities, family, housework). Moderate pain or discomfort. Extremely anxious or depressed.		During the FIRST YEAR,	No problems in walking about. No problems with self-care. No problems with performing your usual activities (e.g.work, study, leisure activities, family, housework). No pain or discomfort. Not anxious or depressed.
Then, during the SECOND YEAR	Some problems in walking about. Some problems with self-care. Some problems with performing your usual activities (e.g.work, study, leisure activities, family, housework). Moderate pain or discomfort. Moderately anxious or depressed.		Then, during the SECOND YEAR	No problems in walking about. Some problems with self-care. Some problems with performing your usual activities (e.g.work, study, leisure activities, family, housework). Moderate pain or discomfort. Moderately anxious or depressed.
Then, during the THIRD YEAR	No problems in walking about. Some problems with self-care. Some problems with performing your usual activities (e.g.work, study, leisure activities, family, housework). Moderate pain or discomfort. Moderately anxious or depressed.		Then, during the THIRD YEAR	No problems in walking about. Some problems with self-care. Some problems with performing your usual activities (e.g.work, study, leisure activities, family, housework). Moderate pain or discomfort. Extremely anxious or depressed.
Then, during the LAST YEAR of the treatment	No problems in walking about. No problems with self-care. Some problems with performing your usual activities (e.g.work, study, leisure activities, family, housework). Moderate pain or discomfort. Not anxious or depressed.		Then, during the LAST YEAR of the treatment	Confined to bed. Unable to wash and dress yourself. Unable to perform your usual activities (e.g.work, study, leisure activities, family, housework). Extreme pain or discomfort. Extremely anxious or depressed.

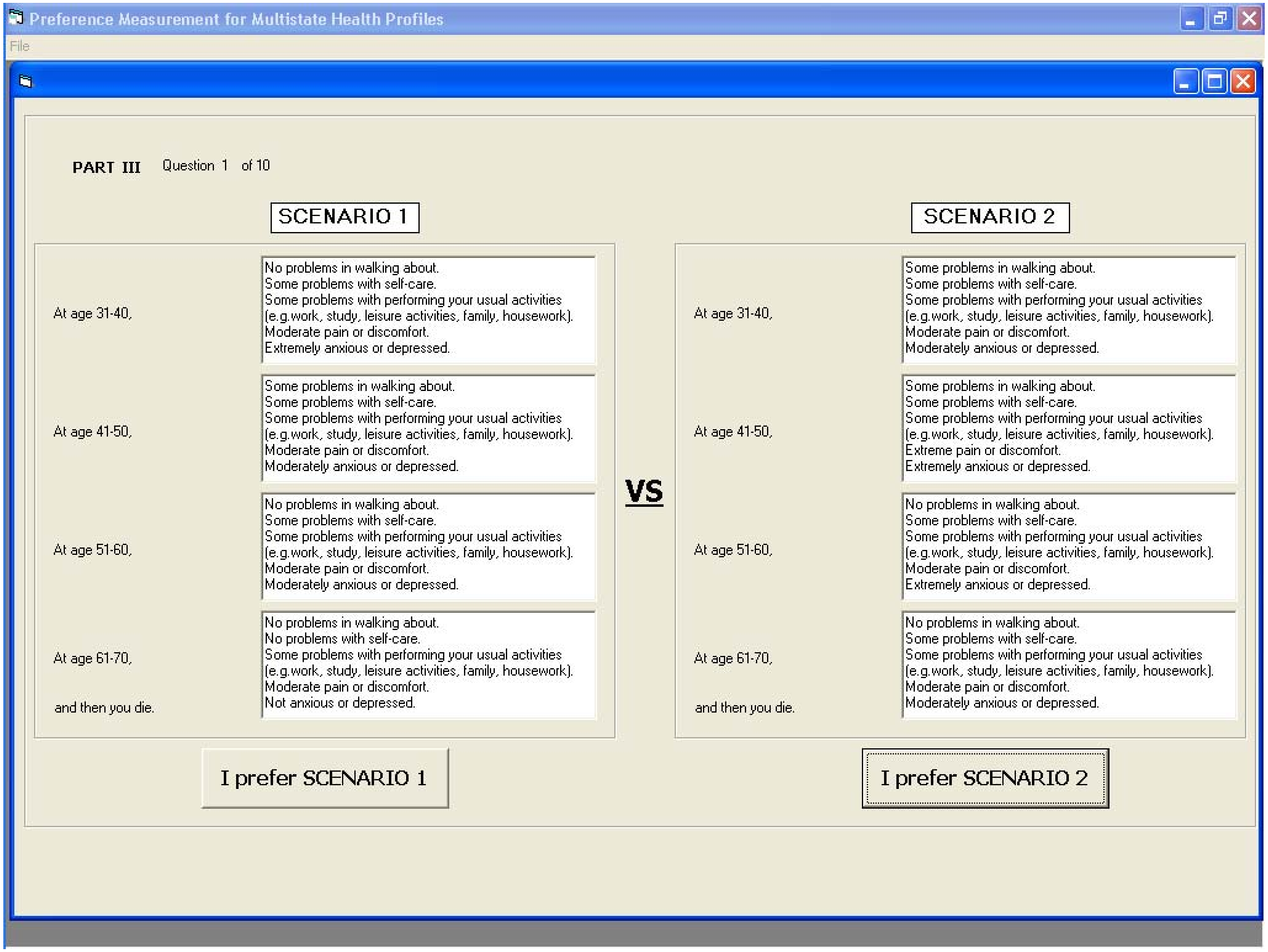
I prefer SCENARIO 1 — I prefer SCENARIO 2


Preference Measurement for Multistate Health Profiles
PART II
Instructions
The paired comparisons you have just done generated your own ranking of the 5 health scenarios from the BEST scenario (1st rank) to the WORST scenario (5th rank). Each health scenario has its own sliding bar. Your task now is to rate how you feel about each health scenario on a scale from 0 to 100 (where 0 is the worst imaginable scenario and 100 is the best imaginable scenario). To do this, you will need to adjust the sliding bar for each health scenario to whichever point on the scale that indicates how you feel about that health scenario.
A description of the health scenario corresponding to each sliding bar will appear in the instructions box as soon as you click on its sliding bar. Realize that the STARTING point of each sliding bar provided to you is NOT your actual rating. It only provides a guidance to your rating based on the rankings generated from the paired comparisons in the previous section. You can go back to these instructions at any time by clicking the SHOW INSTRUCTIONS button on the bottom left corner of this screen.
Remember that the higher value you put on the scale means that the more preferable the health scenario is to you.
Please take your time and adjust these 5 scenarios sliding bars until you feel your rating accurately reflects how you feel. When you are done with the ratings for all health scenarios, please click on the SUBMIT button.
1ST RANK SCENARIO
2ND RANK SCENARIO
3RD RANK SCENARIO
4TH RANK SCENARIO
5TH RANK SCENARIO
100
Best
Worst
0
SHOW INSTRUCTIONS
SUBMIT


Preference Measurement for Multistate Health Profiles
PART II
1ST RANK SCENARIO
During the FIRST YEAR,
No problems in walking about.
Some problems with self-care.
Some problems with performing your usual activities (e.g.work, study, leisure activities, family, housework).
Moderate pain or discomfort.
Moderately anxious or depressed.
Then, during the SECOND YEAR
No problems in walking about.
No problems with self-care.
No problems with performing your usual activities (e.g.work, study, leisure activities, family, housework).
No pain or discomfort.
Not anxious or depressed.
Then, during the THIRD YEAR
No problems in walking about.
Some problems with self-care.
Some problems with performing your usual activities (e.g.work, study, leisure activities, family, housework).
Moderate pain or discomfort.
Extremely anxious or depressed.
Then, during the LAST YEAR of the treatment
Some problems in walking about.
Some problems with self-care.
Some problems with performing your usual activities (e.g.work, study, leisure activities, family, housework).
Extreme pain or discomfort.
Extremely anxious or depressed.
1ST RANK SCENARIO
2ND RANK SCENARIO
3RD RANK SCENARIO
4TH RANK SCENARIO
5TH RANK SCENARIO
100
Best
Worst
0
SHOW INSTRUCTIONS
SUBMIT


Preference Measurement for Multistate Health Profiles
PART II
5TH RANK SCENARIO
During the FIRST YEAR,
No problems in walking about.
No problems with self-care.
Some problems with performing your usual activities (e.g.work, study, leisure activities, family, housework).
Moderate pain or discomfort.
Not anxious or depressed.
Then, during the SECOND YEAR
No problems in walking about.
Some problems with self-care.
Some problems with performing your usual activities (e.g.work, study, leisure activities, family, housework).
Moderate pain or discomfort.
Moderately anxious or depressed.
Then, during the THIRD YEAR
Some problems in walking about.
Some problems with self-care.
Some problems with performing your usual activities (e.g.work, study, leisure activities, family, housework).
Moderate pain or discomfort.
Moderately anxious or depressed.
Then, during the LAST YEAR of the treatment
No problems in walking about.
Some problems with self-care.
Some problems with performing your usual activities (e.g.work, study, leisure activities, family, housework).
Moderate pain or discomfort.
Extremely anxious or depressed.
1ST RANK SCENARIO
2ND RANK SCENARIO
3RD RANK SCENARIO
4TH RANK SCENARIO
5TH RANK SCENARIO
100
Best
Worst
0
SHOW INSTRUCTIONS
SUBMIT


Preference Measurement for Multistate Health Profiles
File
Instructions
PART III
Instructions
In each of the following screens, you will see two hypothetical health scenarios, SCENARIO 1 and SCENARIO 2. Each scenario will ask you to imagine that now you are 30 years old and will live for 40 more years.
Each of the two scenarios on each screen is composed of a sequence of 4 health states over your lifetime. Each health state has a 10-year duration. Please read the description of each health scenario carefully.
On each screen, you will be asked to select the scenario that is preferred over the other one. To do this, you will simply need to click the corresponding button at the bottom of the screen.
Click on NEXT button to start.
NEXT


Preference Measurement for Multistate Health Profiles

File

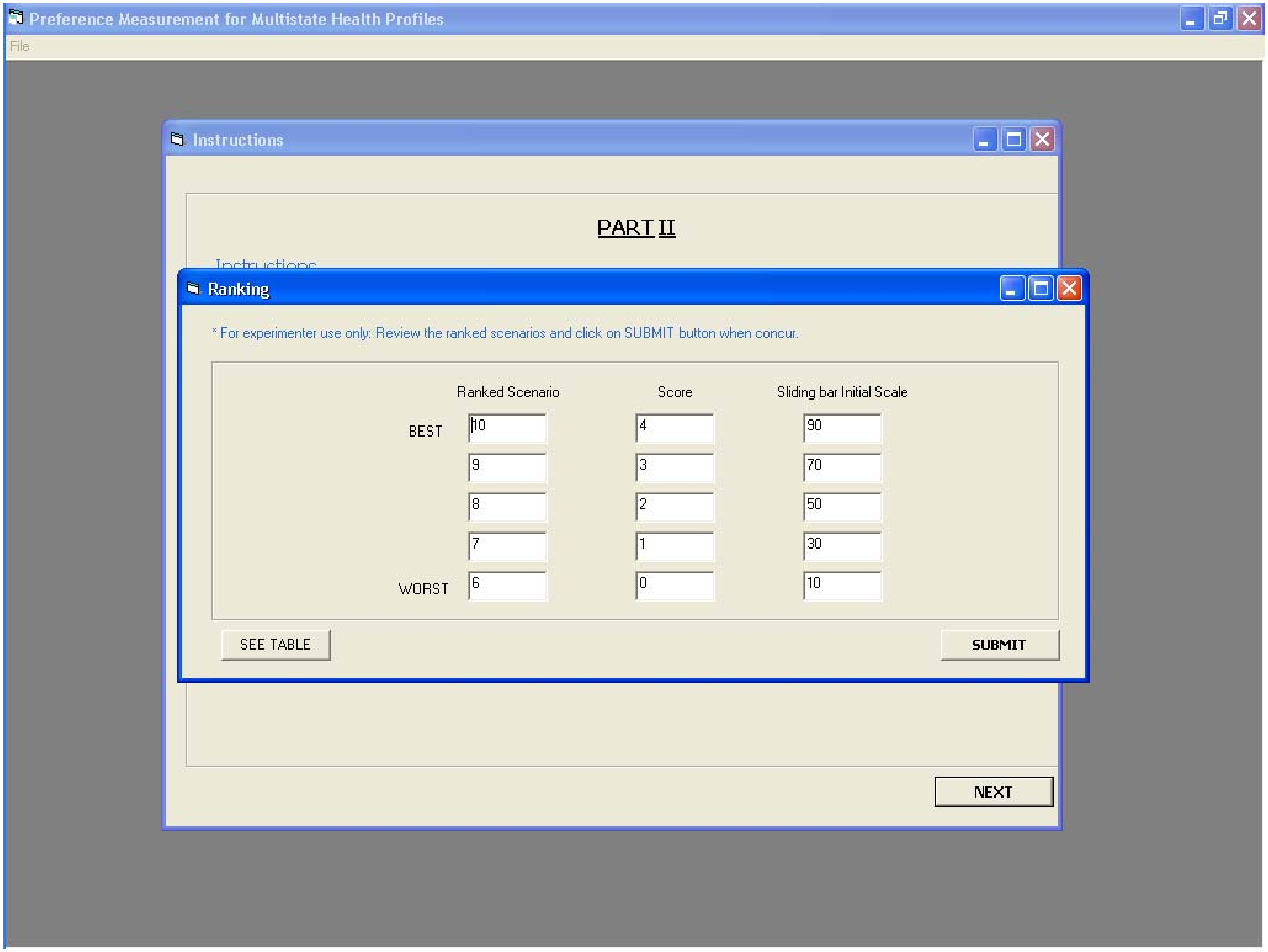
PART III Question 1 of 10

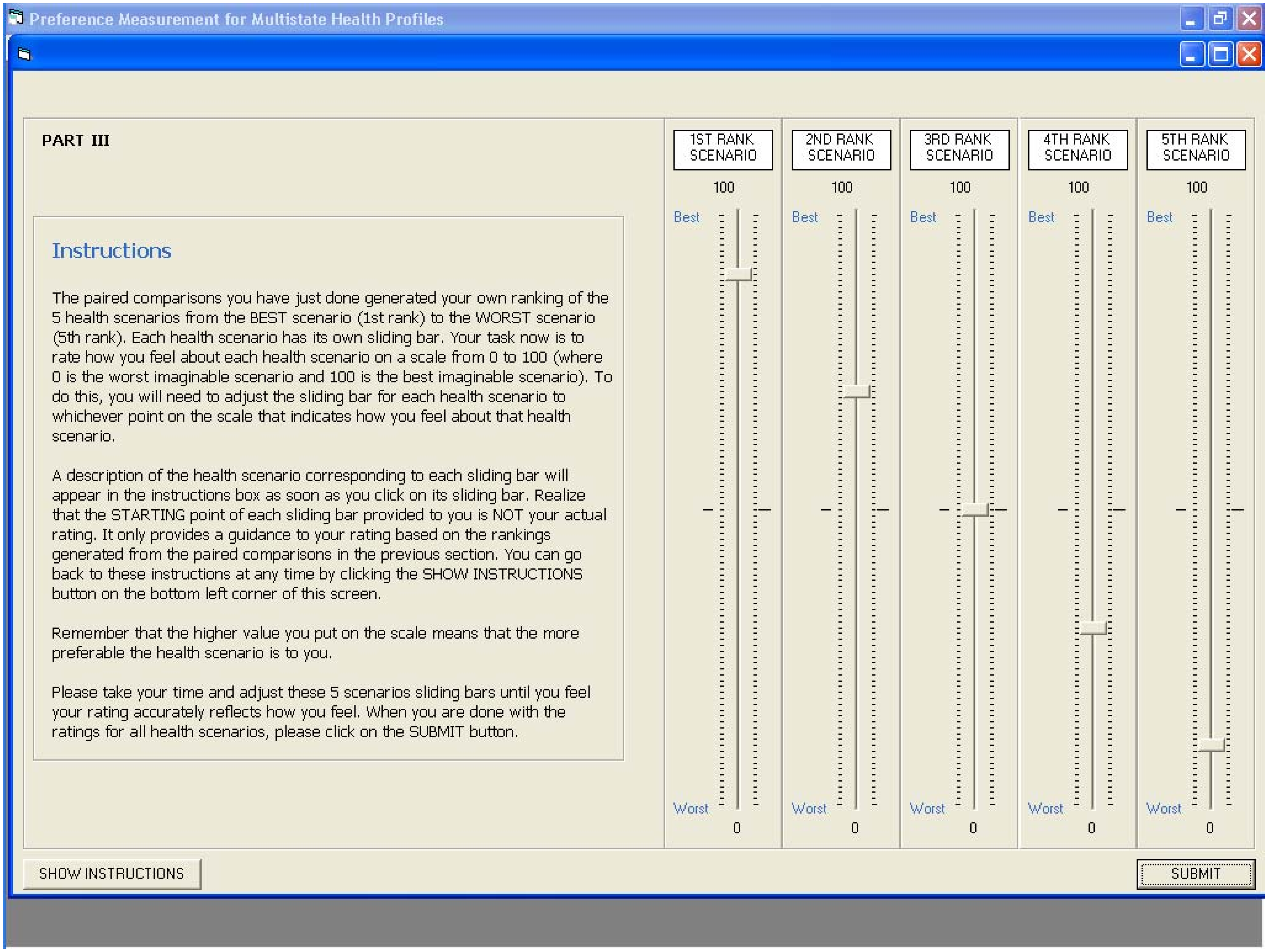
SCENARIO 1		SCENARIO 2	
At age 31-40,	No problems in walking about. Some problems with self-care. Some problems with performing your usual activities (e.g.work, study, leisure activities, family, housework). Moderate pain or discomfort. Extremely anxious or depressed.	At age 31-40,	Some problems in walking about. Some problems with self-care. Some problems with performing your usual activities (e.g.work, study, leisure activities, family, housework). Moderate pain or discomfort. Moderately anxious or depressed.
At age 41-50,	Some problems in walking about. Some problems with self-care. Some problems with performing your usual activities (e.g.work, study, leisure activities, family, housework). Moderate pain or discomfort. Moderately anxious or depressed.	At age 41-50,	Some problems in walking about. Some problems with self-care. Some problems with performing your usual activities (e.g.work, study, leisure activities, family, housework). Extreme pain or discomfort. Extremely anxious or depressed.
At age 51-60,	No problems in walking about. Some problems with self-care. Some problems with performing your usual activities (e.g.work, study, leisure activities, family, housework). Moderate pain or discomfort. Moderately anxious or depressed.	At age 51-60,	No problems in walking about. Some problems with self-care. Some problems with performing your usual activities (e.g.work, study, leisure activities, family, housework). Moderate pain or discomfort. Extremely anxious or depressed.
At age 61-70, and then you die.	No problems in walking about. No problems with self-care. Some problems with performing your usual activities (e.g.work, study, leisure activities, family, housework). Moderate pain or discomfort. Not anxious or depressed.	At age 61-70, and then you die.	No problems in walking about. Some problems with self-care. Some problems with performing your usual activities (e.g.work, study, leisure activities, family, housework). Moderate pain or discomfort. Moderately anxious or depressed.

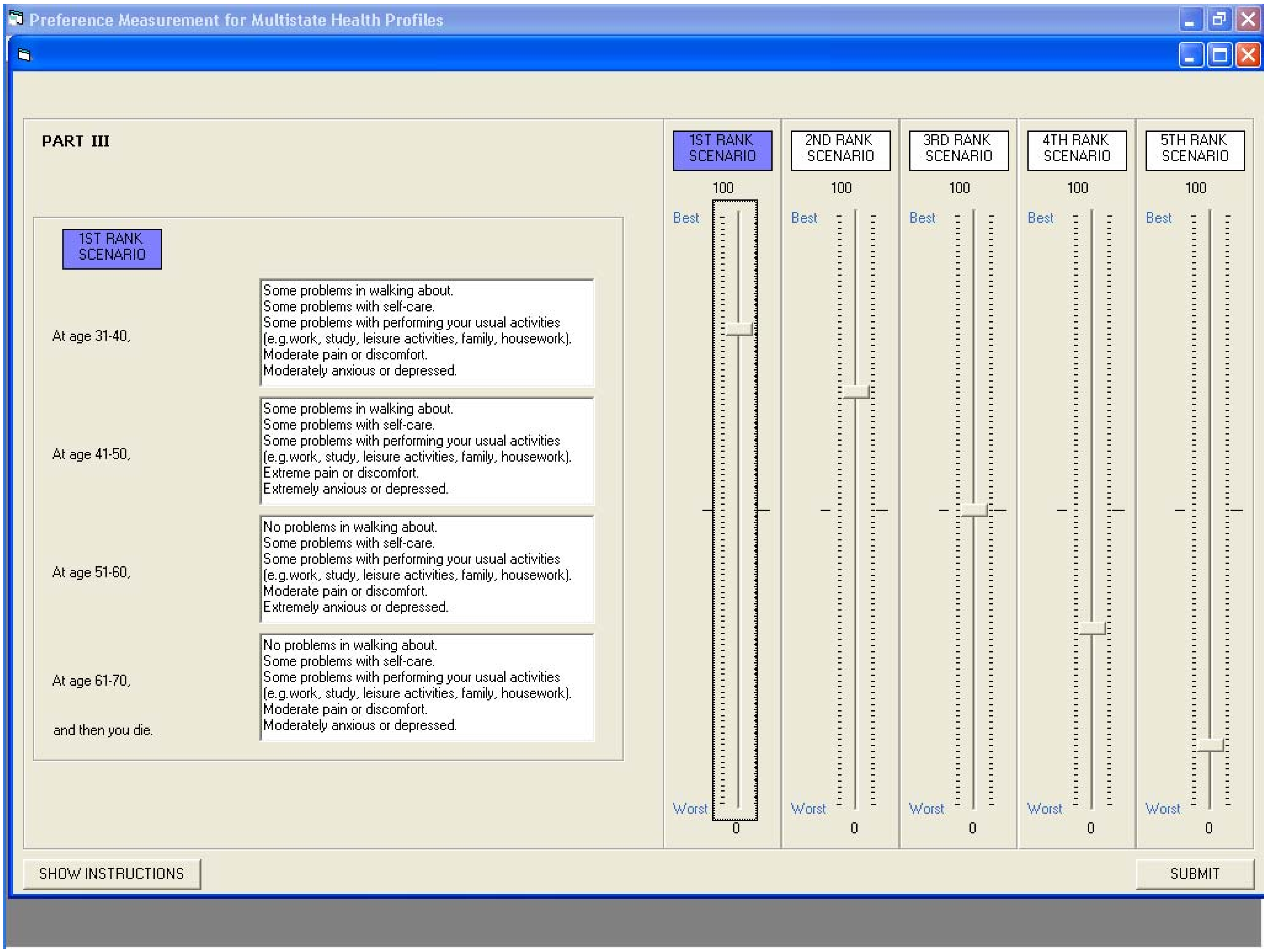
VS

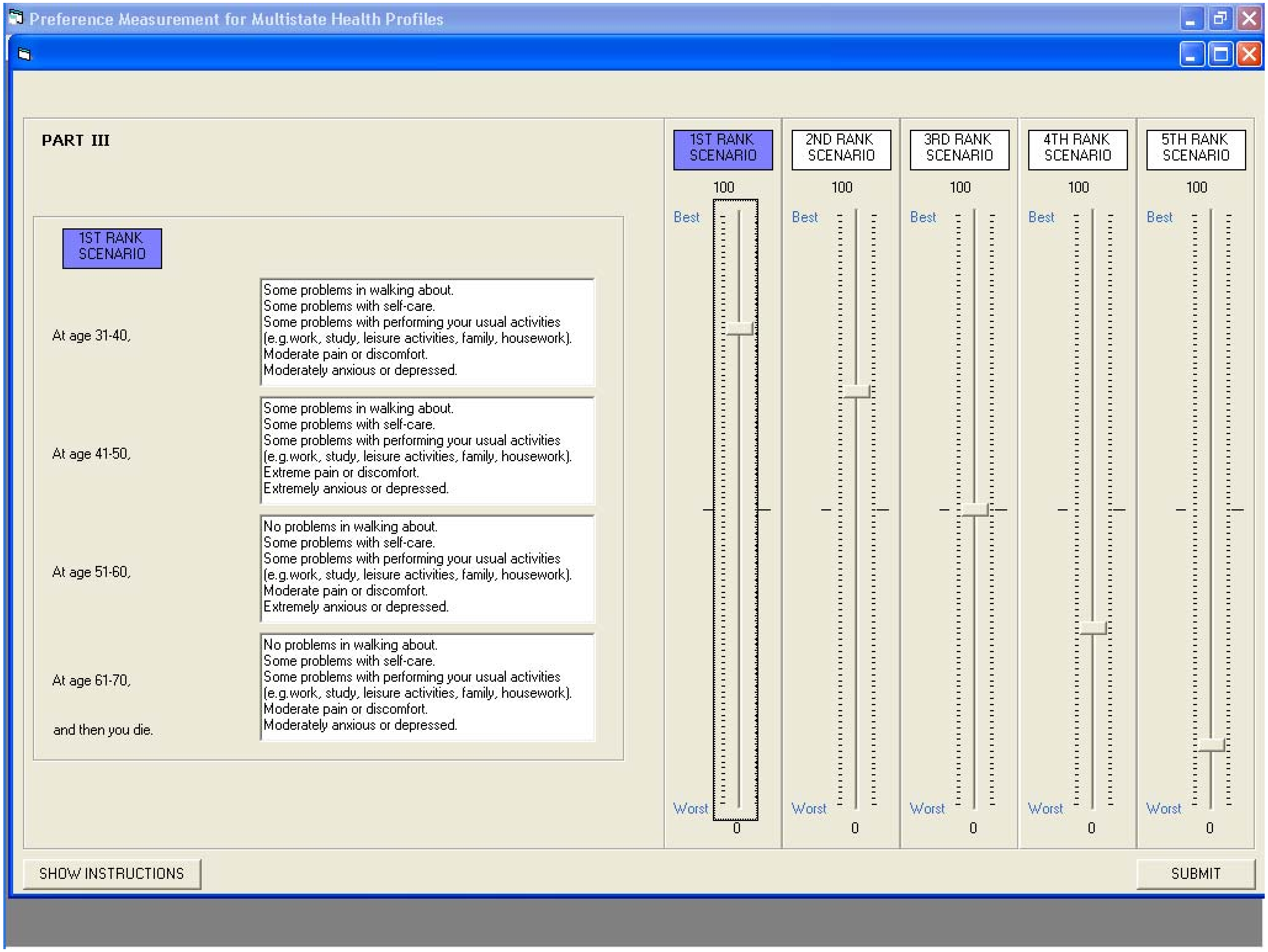
I prefer SCENARIO 1

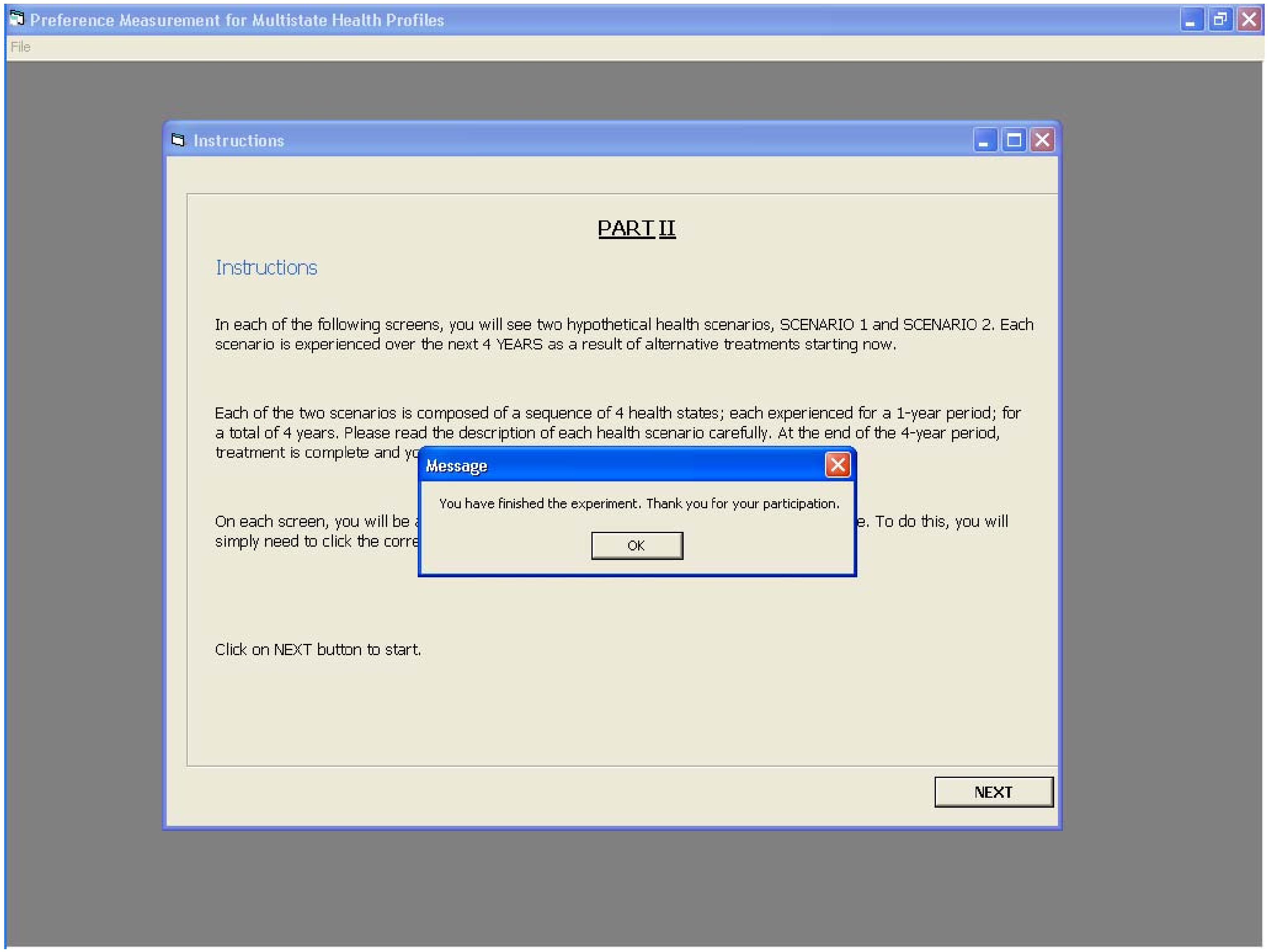
I prefer SCENARIO 2


Preference Measurement for Multistate Health Profiles
File
Instructions
PART II
Ranking
* For experimenter use only: Review the ranked scenarios and click on SUBMIT button when concur.
Ranked Scenario
Score
Sliding bar Initial Scale
BEST
10
4
90
9
3
70
8
2
50
7
1
30
WORST
6
0
10
SEE TABLE
SUBMIT
NEXT


Preference Measurement for Multistate Health Profiles
PART III
Instructions
The paired comparisons you have just done generated your own ranking of the 5 health scenarios from the BEST scenario (1st rank) to the WORST scenario (5th rank). Each health scenario has its own sliding bar. Your task now is to rate how you feel about each health scenario on a scale from 0 to 100 (where 0 is the worst imaginable scenario and 100 is the best imaginable scenario). To do this, you will need to adjust the sliding bar for each health scenario to whichever point on the scale that indicates how you feel about that health scenario.
A description of the health scenario corresponding to each sliding bar will appear in the instructions box as soon as you click on its sliding bar. Realize that the STARTING point of each sliding bar provided to you is NOT your actual rating. It only provides a guidance to your rating based on the rankings generated from the paired comparisons in the previous section. You can go back to these instructions at any time by clicking the SHOW INSTRUCTIONS button on the bottom left corner of this screen.
Remember that the higher value you put on the scale means that the more preferable the health scenario is to you.
Please take your time and adjust these 5 scenarios sliding bars until you feel your rating accurately reflects how you feel. When you are done with the ratings for all health scenarios, please click on the SUBMIT button.
1ST RANK SCENARIO
2ND RANK SCENARIO
3RD RANK SCENARIO
4TH RANK SCENARIO
5TH RANK SCENARIO
100
Best
Worst
0
SHOW INSTRUCTIONS
SUBMIT


Preference Measurement for Multistate Health Profiles
PART III
1ST RANK SCENARIO
At age 31-40,
Some problems in walking about.
Some problems with self-care.
Some problems with performing your usual activities
(e.g.work, study, leisure activities, family, housework).
Moderate pain or discomfort.
Moderately anxious or depressed.
At age 41-50,
Some problems in walking about.
Some problems with self-care.
Some problems with performing your usual activities
(e.g.work, study, leisure activities, family, housework).
Extreme pain or discomfort.
Extremely anxious or depressed.
At age 51-60,
No problems in walking about.
Some problems with self-care.
Some problems with performing your usual activities
(e.g.work, study, leisure activities, family, housework).
Moderate pain or discomfort.
Extremely anxious or depressed.
At age 61-70,
No problems in walking about.
Some problems with self-care.
Some problems with performing your usual activities
(e.g.work, study, leisure activities, family, housework).
Moderate pain or discomfort.
Moderately anxious or depressed.
and then you die.
1ST RANK SCENARIO
2ND RANK SCENARIO
3RD RANK SCENARIO
4TH RANK SCENARIO
5TH RANK SCENARIO
100
Best
Worst
0
SHOW INSTRUCTIONS
SUBMIT


Preference Measurement for Multistate Health Profiles
PART III
1ST RANK SCENARIO
At age 31-40,
Some problems in walking about.
Some problems with self-care.
Some problems with performing your usual activities
(e.g.work, study, leisure activities, family, housework).
Moderate pain or discomfort.
Moderately anxious or depressed.
At age 41-50,
Some problems in walking about.
Some problems with self-care.
Some problems with performing your usual activities
(e.g.work, study, leisure activities, family, housework).
Extreme pain or discomfort.
Extremely anxious or depressed.
At age 51-60,
No problems in walking about.
Some problems with self-care.
Some problems with performing your usual activities
(e.g.work, study, leisure activities, family, housework).
Moderate pain or discomfort.
Extremely anxious or depressed.
At age 61-70,
and then you die.
No problems in walking about.
Some problems with self-care.
Some problems with performing your usual activities
(e.g.work, study, leisure activities, family, housework).
Moderate pain or discomfort.
Moderately anxious or depressed.
1ST RANK SCENARIO
2ND RANK SCENARIO
3RD RANK SCENARIO
4TH RANK SCENARIO
5TH RANK SCENARIO
100
Best
Worst
0
SHOW INSTRUCTIONS
SUBMIT



Preference Measurement for Multistate Health Profiles
File
Instructions
PART II
Instructions
In each of the following screens, you will see two hypothetical health scenarios, SCENARIO 1 and SCENARIO 2. Each scenario is experienced over the next 4 YEARS as a result of alternative treatments starting now.
Each of the two scenarios is composed of a sequence of 4 health states; each experienced for a 1-year period; for a total of 4 years. Please read the description of each health scenario carefully. At the end of the 4-year period, treatment is complete and yo
On each screen, you will be a
e. To do this, you will
simply need to click the corre
Click on NEXT button to start.
NEXT
Message
You have finished the experiment. Thank you for your participation.
OK

APPENDIX B

OUTLIER ANALYSIS (PHASE 1)

Data from the subjects was stored in the computer when the subjects performed the tasks. Outlier needed to be identified before performing any analyses. After scores were linearly normalized for each subject by mapping best health state (perfect health) and worse health state (death), specified in the stage of unconditional preference assessments, to 1.00 and 0.00, respectively. Box plots were used as a tool in identifying outliers. Responses with values more than three box lengths (interquartile range) from the upper or lower edge of the box were identified as extremes and the entire rows of the data were removed. Moreover, cases that contained normalized scores that were higher than 1.00 or lower than 0.00 were removed. Table B.1 summarizes demographics characteristics of the subjects that were removed from the analyses. Figures B.1, B.2, and B.3 illustrate boxplots for conditional preference scores for Design A, B, and C.

Table B.1: Demographics of outliers

Subject #	**7**	**14**	**21**	**30**	**32**	**81**	**97**
Gender							
Female		X	X	X	X		
Male	X					X	X
Ethnic Background							
White, not of Hispanic origin		X		X	X		
Black, not of Hispanic origin			X				
Hispanic							
Asian or Pacific Islander	X					X	X
American Indian or Alaskan Native							
Others							
Age (years)	22	21	21	21	22	19	26
Major field of study							
Industrial Engineering	X	X	X	X	X	X	X
Other - Non-Engineering							
Other - Engineering							
Degree of study							
B.S.	X	X	X	X		X	
M.S.					X		
Ph.D.							X
Other							
Health problem in yourself							
Yes, major issue(s)					X		
Yes, minor issue(s)				X		X	X
No	X	X	X				
Health problem in your family							
Yes, major issue(s)					X	X	
Yes, minor issue(s)		X	X	X			X
No	X						
Health problem in someone else close to you							
Yes, major issue(s)					X		
Yes, minor issue(s)			X	X			X
No	X	X				X	
General health condition							
Excellent					X		
Very Good	X	X	X	X			X
Good						X	
Fair							
Poor							

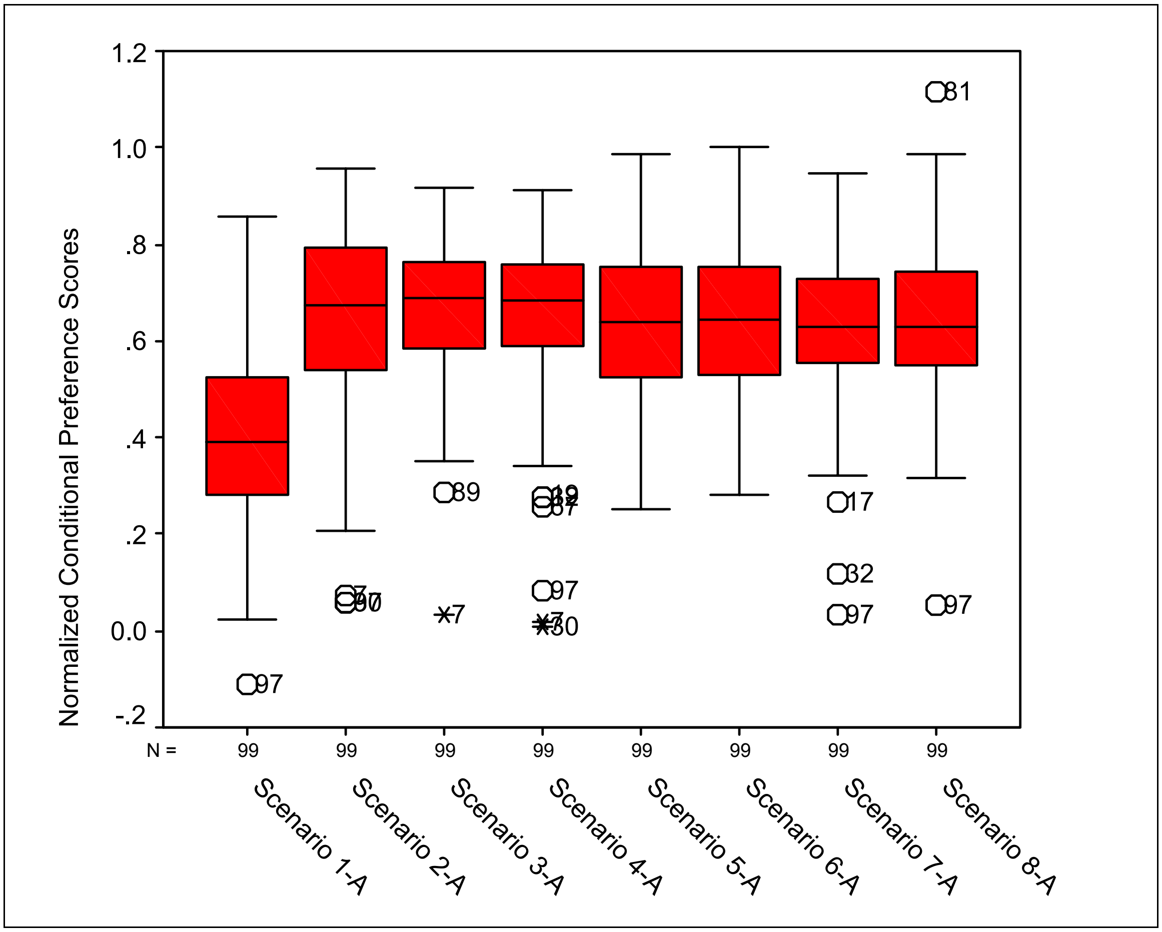


Figure B.1: Boxplots for conditional preference scores in Design A (* indicates an extreme; ○ indicates an outlier)

From Figure B.1, subject #97 was removed due to lower score than 0.00 in scenario 1-A (-0.113) and subject #81 were removed due to higher than 1.00 score in scenario 8-A (1.116). Moreover, subject #7 and subject#30 were removed since they were identified as extremes in scenarios 3-A and 4-A.

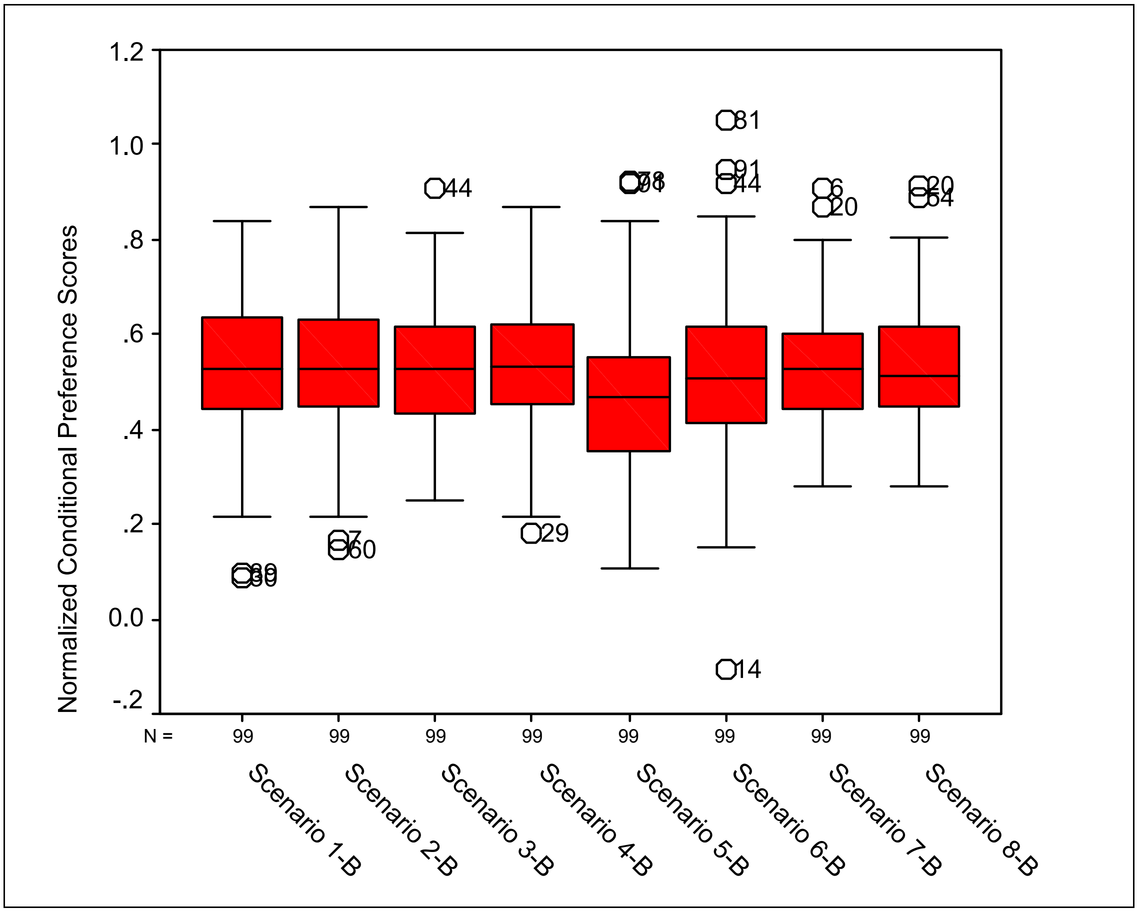


Figure B.2: Boxplots for conditional preference scores in Design B (* indicates an extreme; ○ indicates an outlier)

From Figure B.2, subject #14 was removed due to lower score than 0.00 in scenario 6-B (-0.105) and subject #81 were removed due to higher than 1.00 score in scenario 6-B (1.053).

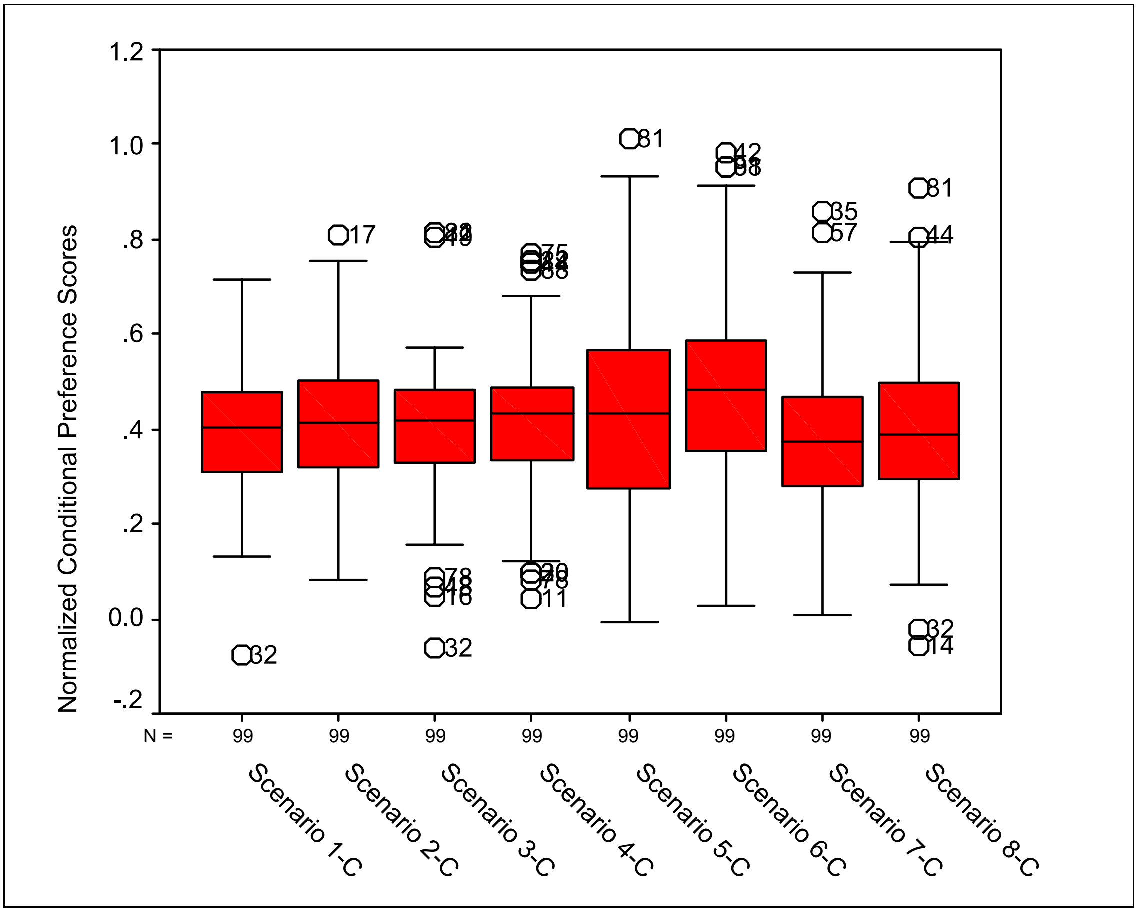


Figure B.3: Boxplots for conditional preference scores in Design C (* indicates an extreme; ○ indicates an outlier)

From Figure B.3, subject #32 was removed due to lower score than 0.00 in scenario 1-C (-0.079), scenario 3-C (-0.059), and scenario 8-C (-0.020). Subject #81 was removed due to higher than 1.00 score in scenario 5-C (1.011). Subject #14 was removed due to the lower than 0.00 score in scenario 8-C (-0.055). Additionally, subject #21 was removed due to the lower than 0.00 score in scenario 5-C (-0.001).

APPENDIX C

OUTLIER ANALYSIS (PHASE 2)

In addition to the outlier analysis in Phase 1, data that was used in Phase 2 included conditional preference scores from an additional 7 scenarios, unconditional preference scores for 7 health states that were selected for the experiment, and holistic preference scores for the health profiles. After scores were linearly normalized for each subject by mapping best health state (perfect health) and worse health state (death), specified in the stage of unconditional preference assessments, to 1.00 and 0.00, respectively. Box plots were used as a tool in identifying outliers. Responses with values more than three box lengths (interquartile range) from the upper or lower edge of the box were identified as extremes and the entire rows of the data were removed. Moreover, cases that contained normalized scores that were higher than 1.00 or lower than 0.00 were removed. Table C.1 summarizes demographics characteristics of the subjects that were removed from the analyses in addition to those in Phase 1. Figures C.1, C.2, and C.3 illustrate boxplots that were used in outlier analysis in Phase 2

Table C.1: Demographics of outliers in addition to those identified in Phase 1

Subject #	**11**	**53**	**67**	**70**	**87**
Gender					
Female	X		X		X
Male		X		X	
Ethnic Background					
White, not of Hispanic origin	X				
Black, not of Hispanic origin					
Hispanic		X			
Asian or Pacific Islander			X	X	X
American Indian or Alaskan Native					
Others					
Age (years)	20	20	20	25	26
Major field of study					
Industrial Engineering	X		X		X
Other - Non-Engineering					
Other - Engineering				X	
Degree of study					
B.S.	X	X	X	X	
M.S.					
Ph.D.					X
Other					
Health problem in yourself					
Yes, major issue(s)				X	
Yes, minor issue(s)	X		X		
No		X			X
Health problem in your family					
Yes, major issue(s)			X	X	
Yes, minor issue(s)	X	X			X
No					
Health problem in someone else close to you					
Yes, major issue(s)		X	X		
Yes, minor issue(s)	X				X
No				X	
General health condition					
Excellent	X	X			
Very Good			X	X	
Good					X
Fair					
Poor					

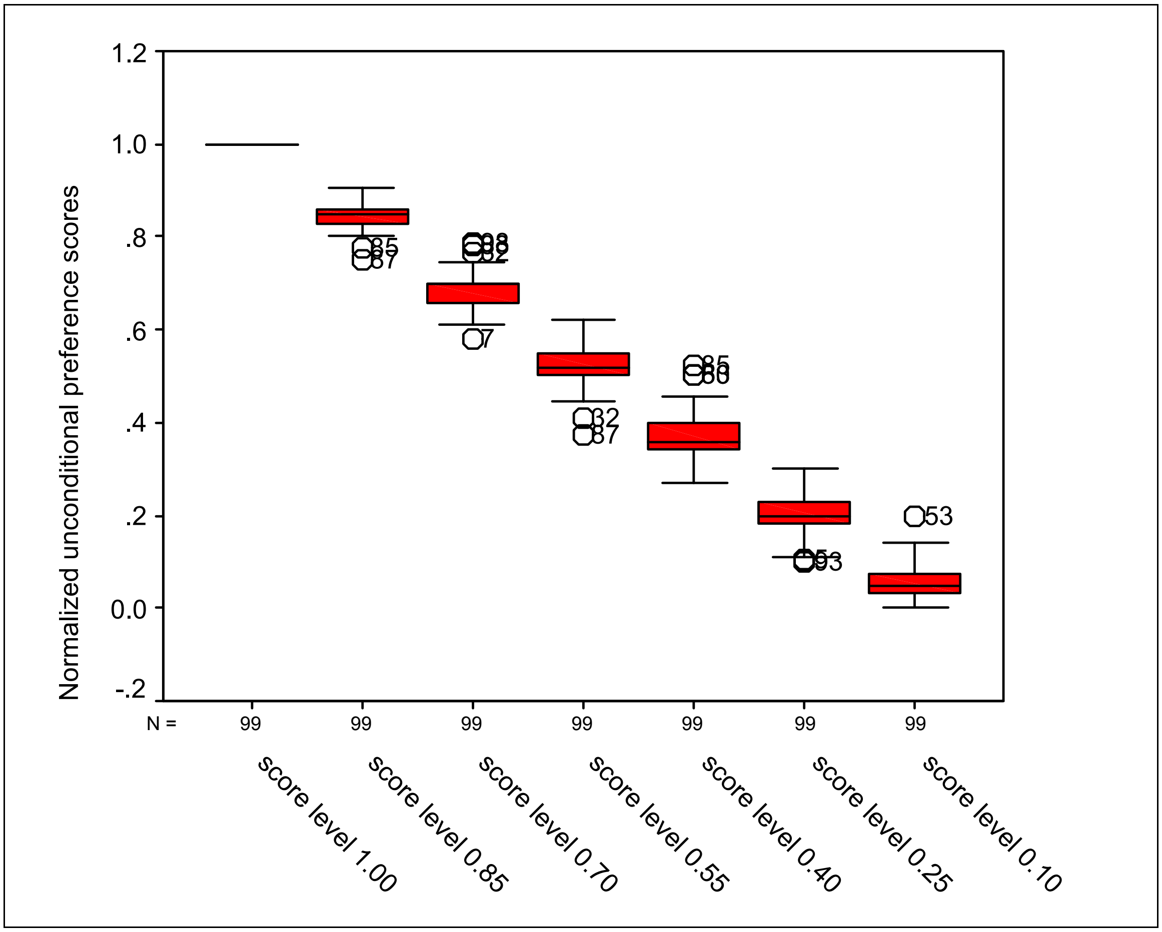


Figure C.1: Boxplots for unconditional preference scores for health states that were selected for the experiment (* indicates an extreme; ○ indicates an outlier)

From Figure C.1, none of the data was identified as extremes and removed.

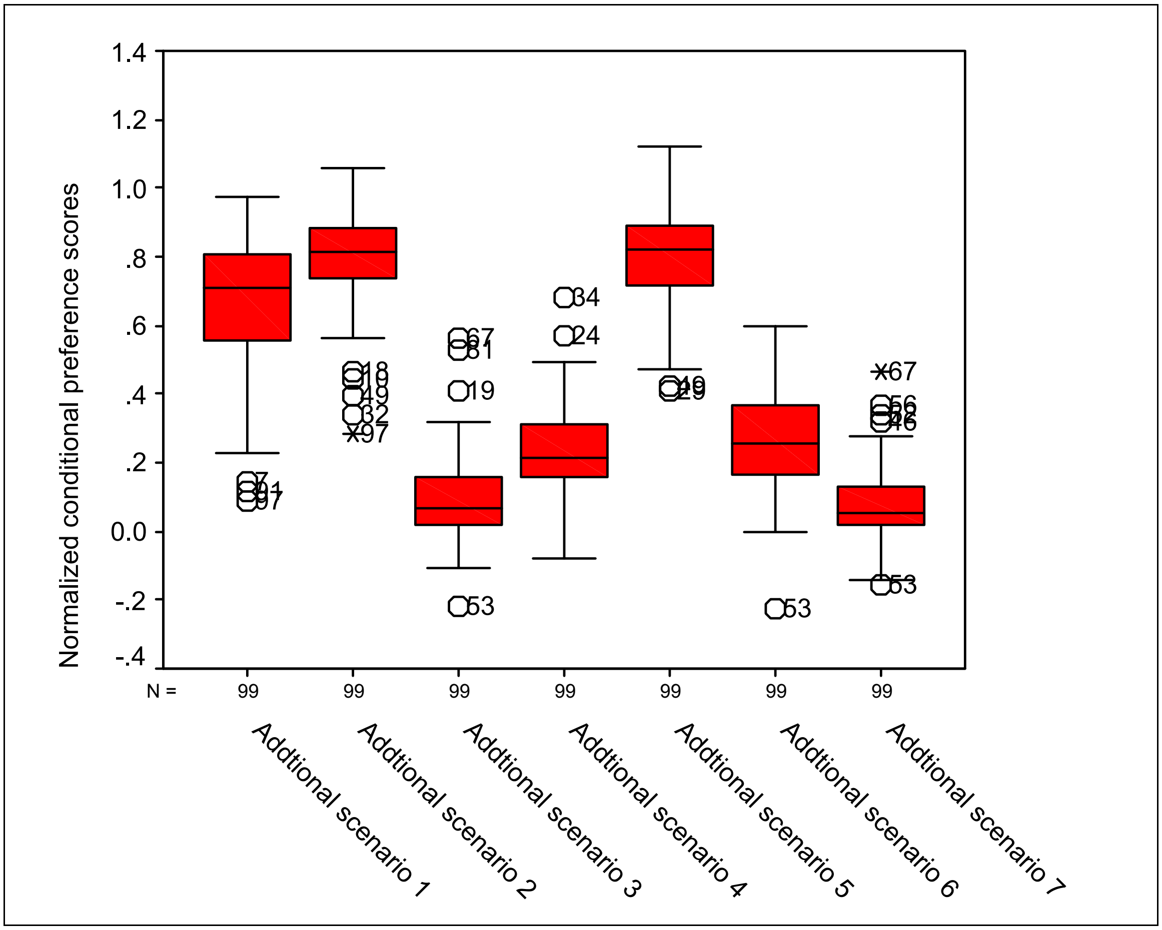


Figure C.2: Boxplots for conditional preference scores in the additional 7 scenarios (* indicates an extreme; ○ indicates an outlier)

From Figure C.2, subject #67 was removed since it was identified as an extreme in scenario 7. Note that subject #97 was also identified as an extreme (scenario 2) but was already removed in Phase 1. Subject #53 was removed since it had scores lower than 0.00 in scenario 3 (-0.218), scenario 6 (-0.227) and scenario 8 (-0.155). Subject # 87 was removed due to a score lower than 0.00 in scenario 8 (-0.003). Subject # 70 was removed due to a score higher than 1.00 in scenario 2 (1.061). Lastly, Subject #11 was removed due to a score higher than 1.00 in scenario 5 (1.062).

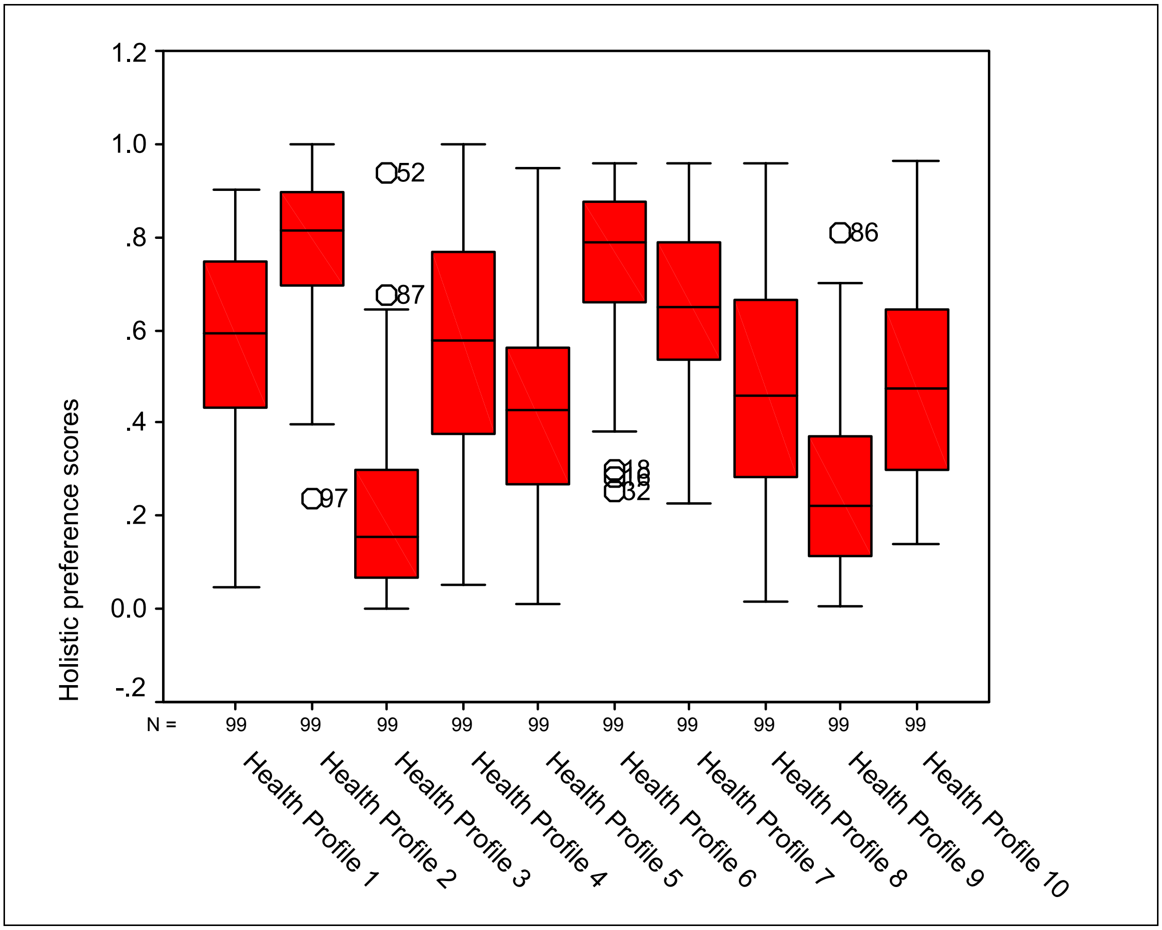


Figure C.3: Boxplots for holistic preference scores (* indicates an extreme; ○ indicates an outlier)

From Figure C.3, none of the data was identified as extremes and removed.

REFERENCES

Ariely D. (1998). "Combining Experiences Over Time: The Effects of Duration, Intensity Changes and On-Line Measurements on Retrospective Pain Evaluations." Journal of Behavioral Decision Making 11: 19-45.

Ariely D. and Carmon Z. (2000). "Gestalt characteristics of experiences: The defining features of summarized events." Journal of Behavioral Decision Making 13: 191-201.

Ariely D. and Zauberman G. (2000). "On the making of an experience: The effects of breaking and combining experiences on their overall evaluation." Journal of Behavioral Decision Making 13: 219-232.

Ariely D. and Loewenstein G. (2000). "When does duration matter in judgment and decision making?" Journal of Experimental Psychology: General 129(4): 508-523.

Baumgartner H., Sujan M. and Padgett D. (1997). "Patterns of affective reactions to advertisements: the integration of moment-to-moment responses into overall judgments." Journal of Marketing Research 34: 219-232.

Benzion U., Rapoport A. and Yagil J. (1989). "Discount rates inferred from decisions: An experimental study." Management Science 35: 270-284.

Berndsen M. and van der Pligt J. (2001). "Time is on my side: optimism in intertemporal choice." Acta Psychologica 108: 173-186.

Bleichrodt H. (1995). "QALYs and HYEs: Under what conditions are they equivalent?" Journal of Health Economics 14: 17-37.

Bleichrodt H. and Johannesson M. (1996). "The Validity of QALYs: An Experimental Test of Constant Proportional Tradeoff and Utility Independence." Medical Decision Making 17: 21-32.

Bleichrodt H., Wakker P. and Johannesson M. (1997). "Characterizing QALYs by risk neutrality." Journal of Risk and Uncertainty 15: 107-114.

Bleichrodt H. and Johannesson M. (2001). "Time preference for health: A test of stationarity versus decreasing timing aversion." Journal of Mathematical Psychology 45: 265-282.

Busschbach J. J. V., McDonnell J. and van Hout B. A. (1997). Testing different parametric relations between the EuroQol Health description and health valuation in students. EuroQol Plenary Meeting Conference Proceedings 1996, Oslo: National Institute of Public Health.

Busschbach J. J. V., McDonnell J., Essink-Bot M. L. and van Hout B. A. (1999). "Estimating parametric relationships between health description and health valuation with an application to the EuroQol EQ-5D." Journal of Health Economics 18: 551-571.

Cairns J. (1992). "Health, wealth, and time preference." Project Appraisal 7(1): 31-40.

Chapman G. B. and Elstein A. S. (1995). "Valuing the Future: Temporal discounting of Health and Money." Medical Decision Making 15: 373-88.

Chapman G. B. (1996a). "Expectations and Preferences for Sequences of Health and Money." Organizational Behavior and Human Decision Processes 67(1): 59-75.

Chapman G. B. (1996b). "Temporal discounting and utility for health and money." Journal of Experimental Psychology: Learning, Memory and Cognition 22(3): 771-791.

Chapman G. B. (1998). "Sooner or later: The psychology of intertemporal choice." The psychology of learning and motivation 38: 83-113.

Chapman G. B. and Coups E. J. (1999). "Time Preferences and Preventive Health Behavior: Acceptance of the Influenza Vaccine." Medical Decision Making 19: 307-14.

Chapman G. B., Nelson R. and Hier D. B. (1999). "Familiarity and time preferences: Decision making about treatments for migraine headaches and Crohn's disease." Journal of Experimental Psychology: Applied 5(1): 17-34.

Chapman G. B. (2000). "Preferences for Improving and Declining Sequences of Health Outcomes." Journal of Behavioral Decision Making 13: 203-218.

Chapman G. B., Brewer N. T., Coups E. J., Brownlee S., Leventhal H. and Leventhal E. A. (2001). "Value of the future and preventive health behavior." Journal of Experimental Psychology: Applied 7(3): 235-250.

Chapman G. B. (2002). "Your Money or Your Health: Time Preferences and Trading Money for Health." Medical Decision Making 22: 410-416.

Chapman G. B. (2003). Time discounting of health outcomes. Time and Decision. G. Loewenstein, D. Read and R. F. Baumeister. New York, Russell Sage Foundation: 395-417.

Christensen-Szalanski J. J. (1984). "Discount functions and the measurement of patients' values: women's decisions during childbirth." Medical Decision Making 4: 45-58.

Cropper M. L., Aydede S. K. and Portney P. R. (1992). "Rates of Time Preference for Saving Lives." The American Economic Review 82(2): 469-472.

Diener E., Wirtz D. and Oishi S. (2001). "End effects of rated life quality: The James Dean effect." Psychological Science 12(2): 124-128.

Dolan P. and Gudex C. (1995). "Time preference, duration and health state valuations." Health Economics 4: 289-99.

Dolan P. (1997). "Modeling valuations for EuroQol health states." Medical Care 35: 1095-1108.

Essink-Bot M. L., Stouthard M. E. A. and Bonsel G. J. (1993). "Generalizability of valuations on healht states collected with the EuroQol questionnaire." Health Economics 2: 237-246.

The EuroQol Group (1990). "EuroQol - a new facility for the measurement of health-related quality of life." Health Policy 16: 199-208.

Fanshel S. and Bush J. W. (1970). "A health-status index and its application to health-services outcomes." Operations Research 18: 1021-1066.

Fredrickson B. L. and Kahneman D. (1993). "Duration neglect in retrospective evaluations of affective episodes." Journal of Personality and Social Psychology 65(1): 45-55.

Fuchs V. R. (1982). Time preference and health: an exploratory study. Economic Aspects of Health. V. R. Fuchs. Chicago, University of Chicago Press: 93-120.

Gafni A. (1995). "Time in Health: Can we measure individuals' "Pure Time Preferences"?" Medical Decision Making 15: 31-7.

Ganiats T. G., Carson R. T., Hamm R. M., Cantor S. B., Sumner W., Spann S. J., Hagen M. D. and Miller C. (2000). "Population-based time preferences for future health outcomes." Medical Decision Making 20: 263-270.

Hadorn D. C. (1991). "The role of public values insetting health care priorities." Social Science in Medicine 32: 773-381.

Helson H. (1964). Adaptation-level theory: An experimental and systematic approach to behavior. New York, Harper & Row.

Hsee C. K., Abelson R. P. and Salovey P. (1991). "The relative weighting of position and velocity in satisfaction." Psychological Science 2(4): 263-266.

Hsee C. K. and Abelson R. P. (1991). "Velocity Relation: Satisfaction as a Function of the First Derivative of Outcome Over Time." Journal of Personality and Social Psychology 60(3): 341-7.

Johannesson M. (1994). "QALYs, HYEs and Individual Preferences - A Graphical Illustration." Social Science and Medicine. 39(12): 1623-1632.

Kahneman D. and Tversky A. (1979). "Prospect theory: An analysis of decision under risk." Econometrica 47: 363-391.

Kahneman D., Fredrickson B. L., Schreibner C. A. and Redelmeier D. A. (1993). "When more pain is preferred to less: adding a better end." Psychological Science 4: 401-5.

Kind P., Dolan P., Gudex C. and Willaims A. (1998). "Variations in population health status: results from a United Kingdom national questionnaire survey." British Medical Journal 316: 736-741.

Krabbe P. F. M. and Bonsel G. J. (1998). "Sequence Effects; Health Profiles, and the QALY Model: In Search of Realistic Modeling." Medical Decision Making 18: 178-88.

Kuppermann M., Shiboski S., Feeny D., Elkin E. P. and Washington E. (1997). "Can Preference scores for Discrete States Be Used to Derive. Preference Scores for an Entire Path of Events?" Medical Decision Making 17: 42-55.

Langer T., Sarin R. and Weber M. (2000). The retrospective evaluation of payment sequences: Duration neglect and Peak-and-End-Effects. Working Paper, University of Manheim.

Lipscomb J. (1989). "Time preferences for health in cost-effectiveness analysis." Medical Care 27: s233-s253.

Loewenstein G. F. (1987). "Anticipation and the valuation of delayed consumption." Economic Journal 97: 666-684.

Loewenstein G. F. and Prelec D. (1991). "Negative time preference." American Economic Review 81: 347-52.

Loewenstein G. and Sicherman N. (1991). "Do workers prefer increasing wage profiles?" Journal of Labor Economics 9: 67-84.

Loewenstein G. F. and Prelec D. (1993). "Preferences for Sequences of Outcomes." Psychological Review 100(1): 91-108.

MacKeigan L. D., Larson L. N., Draugalis J. R., Bootman J. L. and Burns L. R. (1993). "Time preference for health gains versus health losses." Pharmacoeconomics 3: 374-386.

Mackeigan L. D., O'Brien B. J. and Oh P. I. (1999). "Holistic versus Composite Preferences for Lifetime Treatment Sequences for Type 2 Diabetes." Medical Decision Making 19: 113-21.

McNeil B. J., Weichselbaum R. and Pauker S. G. (1978). "Fallacy of the Five-Year Survival in Lung Cancer." New England Journal of Medicine 299: 1397-1401.

Mehrez A. and Gafni A. (1987). "An empirical evaluation of two assessment methods for utility measurement for life years." Socio-Economic Planning Sciences 21(371-375).

Miller N. and Campbell D. T. (1959). "Recency and primacy in persuasion as a function of the timing of speeches and measurements." Journal of Abnormal and Social Psychology 59: 1-9.

Miyamoto J. M. and Eraker S. A. (1988). "A multiplicative model of the utility of survival duration and health quality." Journal of Experimental Psychology: General 117(1): 3-20.

Miyamoto J. M., Wakker P., Bleichrodt H. and Peters H. J. M. (1998). "The zero-condition: A simplifying assumption in QALY measurement and multiattribute utility." Management Science 44(6): 839-849.

Myers J. A., McPherson K. M., Taylor W. J., Weatherall M. and McNaughton H. K. (2003). "Duration of condition is unrelated to health-state valuation of the EuroQol." Clinical Rehabilitation 17: 209-215.

Ohinmaa A., Eija H. and Sintonen H. (1996). Modelling EuroQol values of Finnish adult population. EuroQol Plenary Meeting 1995 Discussion Papers, Barcelona: Institut Universitari de Salut Publica de Catalunya.

Pliskin J. S., Shepard D. S. and Weinstein M. C. (1980). "Utility Functions for Life Years and Health Status." Operation Research 28(1): 206-224.

Redelmeier D. A. and Heller D. M. (1993). "Time preference in medical decision making and cost-effectiveness analysis." Medical Decision Making 13: 212-217.

Redelmeier D. A. and Kahneman D. (1996). "Patients' memories of painful medical treatments: real-time and retrospective evaluations of two minimally invasive procedures." Pain 66: 3-8.

Richardson J., Hall J. and Salkeld G. (1996). "The measurement of utility in multiphase health states." International Journal of Technology Assessment in Health Care 12: 151-62.

Ross W. T. and Simonson I. (1991). "Evaluating pairs of experiences: a preference for happy endings." Journal of Behavioral Decision Making 4: 273-82.

Rué M. and Badia X. (1996). The Spanish EuroQol Tariff: Results from the Catalan Health Survey based on self-rated health. EuroQol Plenary Meeting 1995 Discussion Papers, Barcelona: Institut Universitari de Salut Publica de catalunya.

Russell L., Gold M., Siegel J., Daniels N. and Weinstein M. C. (1996). "The role of cost-effectiveness analysis in health and medicine: Panel on cost-effectiveness in health and medicine." Journal of the American Medical Association 276: 1172-1177.

Samuelson P. (1937). "A note on measurement of utility." Review of Economics Studies 4: 155-161.

Schreibner C. A. and Kahneman D. (2000). "Determinants of the remembered utility of aversive sounds." Journal of Experimental Psychology: General 129(1): 27-42.

Shelly M. K. (1994). "Gain/Loss asymmetry in risky intertemporal choice." Organizational Behavior and Human Decision Processes 59(1): 124-159.

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

Stiggelbout A. M., Kiebert G. M., Kievit J., Leer J. W. H., Stoter G. and de Haes J. C. J. M. (1994). "Utility assessment in cancer patients: adjustment of time tradeoff scores for the utility of life years and comparison with standard gamble scores." Medical Decision Making 14: 82-90.

Thaler R. H. (1981). "Some empirical evidence on dynamic inconsistency." Economic Letters 9: 201-207.

Varey C. and Kahneman D. (1992). "Experiences extended across time: evaluation of moments and episodes." Journal of Behavioral Decision Making 5: 169-85.

Verhoef L. C. G., de Haan A. F. J. and van Daal W. A. J. (1994). "Risk attitude in gambles with years of life: Empirical support for prospect theory." Medical Decision Making 14: 194-200.

von Neumann J. and Morgenstern O. (1947). Theory of Games and Economic Behavior. Princeton, NJ, Princeton University Press.

VITA

Thitima Kongnakorn was born on August 15, 1975 in Bangkok, Thailand. She graduated with a Bachelor of Science in Production Engineering in 1996 from King Mongkut's University of Technology Thonburi. She then entered the University of Wisconsin-Madison in the August 1997 to pursue graduate study in the Department of Industrial Engineering with the focus on quality engineering. After receiving her Master of Science in Industrial Engineering in December 1998, she started working on her Ph.D. research in the area of health systems under the direction of Dr. François Sainfort. In January 2001, she transferred to the School of Industrial and Systems Engineering at the Georgia Institute of Technology, in Atlanta, Georgia. There, she continued her work with Dr. Sainfort in health systems and was also co-advised by Dr. Julie A. Jacko in the area of health care informatics.