GUIDE TO DESIGN AND DEVELOPMENT
OF HEALTH-STATE UTILITY
INSTRUMENTATION

William Furlong
David Feeny
George W. Torrance
Ronald Barr
John Horsman

"NOT FOR CITATION WITHOUT PERMISSION"
JUNE, 1990

GUIDE TO DESIGN AND DEVELOPMENT
OF HEALTH-STATE UTILITY

INSTRUMENTATION

William Furlong¹, David Feeny¹,²,⁴,
George W. Torrance¹,²,⁴, Ronald Barr³,
John Horsman⁵

McMaster University
Hamilton, Ontario
L8N 3Z5
TEL: (416) 525-9140
FAX: (416) 546-5211
Telex: 0618347

1. Department of Clinical Epidemiology and Biostatistics
2. Department of Economics
3. Faculty of Business
4. Centre for Health Economics and Policy Analysis
5. Department of Pediatrics
6. Department of Medicine
ABSTRACT


Growing interest in methods of quantifying health-related quality of life has resulted in demand for a manual describing general techniques that may be customized by health researchers and used to measure preferences in applications having a wide variety of objectives. This report is intended to provide practical guidelines to researchers who require information about the design, construction and administration of instruments used to obtain estimates of health utility. The focus is on the steps required to conduct a study in which preferences are measured; the theory that underlies utility measurement is not presented.

Utility scores may be used to assess the health-related quality of life for specific health states. Utility scores are also useful in evaluations, such as cost-utility analyses, in which the costs of health care interventions are compared to their consequences both in terms of quantity and quality of life. In order to obtain reliable, valid, and responsive utility scores, however, the instruments must be carefully designed, tested and executed. Experience to date indicates that careful instrument development requires time, effort and care.

Three major instruments are described: rating scale, standard gamble and time trade-off. Methodologically correct techniques are recommended where possible and suggestions, based on experience, are presented for situations in which theoretical or empirical evidence is lacking. The guide discusses the criteria for selection of techniques, materials, interviewers and respondents. Examples of an interviewer training manual, interview scripts and data recording forms are also included. Detailed diagrams provide specifications for assembling visual aids (i.e., interviewing props) and an extensive bibliography will direct readers interested in specific issues to appropriate references.
TABLE OF CONTENTS

List of Tables ......................................................................................................................................... iii
List of Figures ........................................................................................................................................ iv
Acknowledgements.................................................................................................................................. v

1. Introduction .......................................................................................................................................... 1

2. Methods ............................................................................................................................................... 3

3. Instrumentation .................................................................................................................................... 4

4. Health States ....................................................................................................................................... 7

5. Instrument Testing ............................................................................................................................... 10

6. Interviewers ......................................................................................................................................... 12

7. Respondents ....................................................................................................................................... 14
   7.1 Sample Size, Parameter Estimates and Hypothesis Testing ......................................................... 16

8. Consent.................................................................................................................................................. 23

   9.1 Design and Assembly ...................................................................................................................... 24
   9.2 Interview Schedules and Response Marking Form ........................................................................ 25

10. Chance Board .................................................................................................................................... 26
   10.1 Design and Assembly ................................................................................................................... 28
   10.2 Interview Schedules and Response Booklets ............................................................................. 31
   10.3 Administration ................................................................................................................................ 33

11. Time Trade-Off Board ..................................................................................................................... 35
   11.1 Design and Assembly ................................................................................................................... 36
   11.2 Interview Schedules and Response Booklets ............................................................................. 39
   11.3 Administration ................................................................................................................................ 40

12. Other Visual Aids ............................................................................................................................... 43

13. Conclusions ......................................................................................................................................... 45

Bibliography ........................................................................................................................................... 46

Tables ...................................................................................................................................................... 49

Figures ................................................................................................................................................... 56

Glossary of Terms .................................................................................................................................. 73

APPENDIX I - Example Consent Forms for Parents-of-Patients and General Public Parents - Example ......................................................................................................................... 80

APPENDIX II - Interviewer Manual - Example .................................................................................... 83
TABLE OF CONTENTS CONT'D

APPENDIX III - Feeling Thermometer and Standard Gamble (Chance) Schedule - Example.................................................................101

APPENDIX IV - Response Booklet (Feeling Thermometer, Standard Gamble, Demographic, Interview Assessment sections) - Example....117

APPENDIX V - Cascading Standard Gamble Response Form - Example........129

APPENDIX VI - Time Trade-Off Schedule - Example........................................131

APPENDIX VII - Response Form for Time Trade-Off Questions - Example.....137

APPENDIX VIII - Lists of Materials for Constructing Visual Aids..............139
LIST OF TABLES

1. Examples of Approximate 95% Confidence Intervals for Population Means ................................................................. 50

2. Power Calculation Data for Comparisons of Mean Utility Scores for Independent Experienced (eg., patient) Respondent Groups (eg., males vs. females) ................................................ 51

3. Power Calculation Data for Comparisons of Mean Utility Scores for Independent General Public Respondent Groups (eg., males vs. females) ......................................................... 52

4. Power Calculation Data for Comparisons of Mean Utility Scores for Paired Experienced (eg., patient) Responses (SD = 0.2) ................................................................. 53

5. Power Calculation Data for Comparisons of Mean Utility Scores for Paired General Public Respondents (SD = 0.3) ................................................................. 54

6. Power Calculation Data for Comparisons of Mean Utility Scores for Independent Respondent Groups with Heterogeneity of Variance ................................................................. 55
<table>
<thead>
<tr>
<th>FIGURE</th>
<th>DESCRIPTION</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.0</td>
<td>Flow Diagram of Basic Steps in Development of Health-State Utility Measurements</td>
</tr>
<tr>
<td>1.1</td>
<td>Feeling Thermometer Board Specifications</td>
</tr>
<tr>
<td>1.2(a)</td>
<td>Layout of a Descriptive Card for a Chronic Health State Based on Individual Levels for Seven Attributes</td>
</tr>
<tr>
<td>1.2(b)</td>
<td>Layout of a Descriptive Card for a Short Duration Health State Associated with a Specific Disease and Treatment</td>
</tr>
<tr>
<td>2.1</td>
<td>Chance (Standard Gamble) Board Faceplate Specifications</td>
</tr>
<tr>
<td>2.2</td>
<td>Layout of Chance Board Probabilities Wheel</td>
</tr>
<tr>
<td>2.3</td>
<td>Chance (Standard Gamble) Board Backplate Specifications</td>
</tr>
<tr>
<td>2.4</td>
<td>Flow Diagram of a Conventional Standard Gamble Question</td>
</tr>
<tr>
<td>2.5</td>
<td>Sequence and Placement of Health State Descriptions for Simplified Cascading Exercise</td>
</tr>
<tr>
<td>3.1</td>
<td>Time Trade-Off Board Faceplate Specifications</td>
</tr>
<tr>
<td>3.2</td>
<td>Time Trade-Off Board Backing Plate Specifications</td>
</tr>
<tr>
<td>3.3</td>
<td>Time Trade-Off Board Slider Specifications</td>
</tr>
<tr>
<td>3.4</td>
<td>Examples of Time Trade-Off Board Scales</td>
</tr>
<tr>
<td>3.5</td>
<td>Example of Steps and Formats for Simple Conventional Trade-Off Questions</td>
</tr>
<tr>
<td>4.0</td>
<td>Example of a Card Deck Used to Present Standard Gamble Questions</td>
</tr>
<tr>
<td>4.1</td>
<td>Example of a Chart and Mask Prop Used to Present Time Trade-Off Questions</td>
</tr>
</tbody>
</table>
ACKNOWLEDGEMENTS

Ontario Ministry of Health, Research Grant 01386 (partial funding).

Merck Foundation (partial funding).

Mrs. Lori Scapinello (Word Processing and Typing).

Ms. Mary Helen Blackall (Prototype Instrument Pretesting).

Ms. Marie Townsend (Response Marking Forms and Sample Size).

Mrs. Ann Mohide (Critical Review).

Ms. Roberta Labelle (Critical Review).

Social Data Research Ltd.:

Dr. M. Denton (survey consultant);
Mrs. B. Nussey (survey field supervisor);
Mrs. J. Brown (survey instrument pretesting); and
Mrs. J. Shumacher (survey instrument pretesting).

McMaster University Health Utility Group.

Mrs. Barbara Markham (Other Visual Aids).

Dr. Donald Patrick (Critical Review).

Dr. Dennis A. Revicki (Critical Review).

Mr. John Corea (Critical Review).

Mrs. Carol Siksay (Word Processing and Printing)

Dr. Dev S. Pathak (Critical Review)

Dr. Greg L. Stoddart (Critical Review)

Prof. Alan Williams (Critical Review)

The usual disclaimer applies.
1. INTRODUCTION

Utility is increasingly being used as an outcome measure in the evaluation of health-care services. This reflects the growing interest in the measurement of quality of life. It also reflects the compatibility of utility measurement with economic evaluation measures that include pecuniary (dollar) estimates of the costs and consequences of alternative health-care programs (Feeny, Labelle, and Torrance, 1990).

Froberg and Kane (1989), in a four-part series of papers, review the state-of-the-art in measuring preferences for states of health. Readers who are unfamiliar with the utility approach should read Feeny and Torrance (1989) and Torrance (1987) before reading this guide. This report describes how to design, construct and administer instruments to measure health-state utilities. (Figure 1.0 presents a graphical overview of major topics.) Utilities can be used as a quality-of-life outcome measure. Utilities are required for use in cost-utility analysis, a special form of cost-effectiveness analysis in which the result is expressed in cost per quality-adjusted life-year gained. The analysis and interpretation of utility measurement results are beyond the scope of this guide; interested readers are referred to Drummond et. al. (1987).

Three instruments are described - rating scale, standard gamble and time trade-off. The standard gamble method is the classical method of utility measurement and is based directly on the fundamental axioms of utility theory. Time trade-off is a proxy method which has often been found to give comparable results. The rating scale instrument is very efficient, but unfortunately it does not give cardinal utility measures directly and therefore rating scales are most often
used in conjunction with other methods. (For a description of the conceptual foundations and methodological properties of the three approaches see Torrance 1986, 1987; Feeny and Torrance, 1989).

We have had considerable experience at McMaster University in designing, building and using all three of these instruments, sometimes all in the same study. This report is aimed at researchers who wish to use one or more of these approaches. We present here important information you will need to develop your own instrument. Sections 2 through 8 are relevant to the development of all types of utility instruments. Sections 9 through 12 deal with specific, individual techniques. A glossary of terms is presented in section 16. (It should be noted that the order of presentation of appendices was determined on the basis of a logical flow of tasks in health-state utility measurement rather than by order of citation.) You will note that where possible we recommend the methodologically correct technique, based on theoretical and/or empirical evidence. However, in other instances there are a variety of ways to achieve the same goal and we present recommendations based on our experience. If you need further help, or if you have questions, please contact the authors at McMaster University. You may also wish to consider enrolling in a two-day workshop "Economic Evaluation of Health Care Programs" which we offer twice per year. The workshop provides an introduction to all aspects of economic evaluation, including the measurement of preferences. For details contact: SEMINEX, McMaster University, School of Business, Hamilton, Ontario, Canada L8S 4M4; or phone (416) 525-9140 Ext. 4636, 4105.
2. METHODS

The three basic methods of measuring preferences are:

1) visual analogue rating scale;
2) standard gamble; and
3) time trade-off.

The visual analogue scale is very efficient, but unfortunately it does not give cardinal utility measures directly. However, it is recommended that the rating scale be included in instruments and used in conjunction with another method. The rating scale provides a vehicle for introducing a full set of health states, helps respondents to consider rank preferences of health states, and prepares respondents for subsequent and more difficult steps in the measurement process.

The standard gamble and time trade-off methods provide cardinal utility measures. If possible, the standard gamble is recommended because it is derived from the basic axioms of utility theory as postulated by von Neumann and Morgenstern (1944, 1947) and is therefore valid by definition (i.e., "gold standard" method). However, the time trade-off method has been validated against the standard gamble method (Torrance et. al., 1973 and Torrance, 1976) and was developed specifically for use in health-care evaluations by Torrance et. al. (1972). Time trade-off may be the preferred method for studies that involve subjects who cannot relate to probabilities, or studies that involve comparisons of alternatives in which time trade-off is the major clinically relevant factor (e.g., a treatment that improves quality of life but decreases life expectancy as compared to a treatment that reduces quality of life but does not reduce life expectancy).
3. INSTRUMENTATION

For the purposes of this manual an instrument is defined as all the written, graphical, verbal and non-verbal instructions and equipment required to measure health-state utilities. Although not every measurement task will require extensive instrumentation, it is important to make informed decisions about instrumentation prior to data collection. Informed decisions will improve the accuracy and precision of measurements in addition to promoting efficient collection of data. In order to reduce variability and bias in large surveys that involve many respondents and interviewers, it is important to conduct measurements in a systematic manner. This requires highly standardized instrument components. The following instrument components are, in general, recommended:

1) Health-State Descriptions (see Section 4);
2) Interview Schedule;
3) Response Recording Forms;
4) Visual Aids (Props); and
5) Interviewer Manual (see Section 5 and Appendix II).

Interview schedules and response booklets are highly recommended to ensure that responses are elicited and recorded in a standard manner. Standardization helps increase measurement precision, ease interviewer burden and provides a format for coding in preparation for data entry. The basis of the interview schedule is a written script to be read aloud by the interviewer to the respondent (i.e., interviewee). The script provides all the details and directions that the respondent will require to complete the measurement task. In addition to the script, the schedule should include practical written directions to the interviewer. The interviewer directions must be well defined, visually separated from the script and injected into the appropriate text in a relevant manner. The schedule normally begins with an acknowledgement of the respondent's co-
operation, description of the research group, an assurance of confidentiality and a brief description of the purpose of the interview (see Appendix III).

When the response recording forms are completed they constitute the original data record and are customized to minimize the amount of information that must be written to record accurately interview data. However, interviewers should be encouraged to use these forms to note problems encountered during each interview. These forms may take the form of tables but other formats are often more efficient. For example, some responses may be recorded by check marks in appropriate pre-printed boxes, and in some circumstances parts of the interview schedule may be included in the response recording form booklet (see "family questions" section of Appendix IV).

Although the format and contents of the response recording forms vary according to study objectives, it is usually important to ensure that the cover of each interview response form or booklet clearly identifies the respondent. It is recommended that study identification numbers be used, rather than names, to maintain the confidentiality of responses. The booklet cover should also identify the interviewer and the interview date. In addition, it may be useful to include the time of day at the start and finish of the interview to determine interview scheduling patterns and durations. Scheduling patterns may assist in explaining variations in compliance rates, and interview durations are often necessary for determining interviewer fees or estimating respondent burden.

Visual aids, including graphics, are commonly used to present information in an efficient manner. Visual aids, often referred to as props, must complement and
not detract from the interview schedule. Props are most often employed to provide the respondent with a continuous visual framework representing important issues of direct relevance to the decision-making process or to choices being presented. The props described in this manual were designed to meet the following criteria:

i) pose questions in an unbiased manner to reduce framing effects (Tversky and Kahneman, 1981);

ii) display information clearly to reduce cognitive burden (i.e., effort required to perceive, think and remember) and encourage respondent co-operation;

iii) present an attractive and professional image;

iv) be fully acceptable to trained professional interviewers (i.e., considered to be more helpful than hinderance);

v) be portable (i.e., small, light, self-contained and rugged); and

vi) be inexpensive to construct and easy to assemble using readily available materials.

The importance of the first two criteria cannot be over-emphasized. However, in our experience the most common problem is excessive cognitive burden and therefore all researchers are cautioned to avoid presenting complicated health-state scenarios, confusing graphics, or using unnecessary jargon. If possible most information should be presented verbally, visual aids (i.e., props and written scenarios) should consist of precise simple labels and utility measures should be elicited at the beginning of interviews before less demanding functional status or other questions are administered (Mohide et. al., 1988).

Many different props have been designed but this manual will concentrate on the
three types developed most recently and generally recommended. These props are commonly known as the Feeling Thermometer (see Section 9), Chance Board (see Section 10) and Time Trade-Off Board (see Section 11) which are terms used to describe specific formats of rating scales, standard gamble and time trade-off techniques, respectively. Appendix VIII presents lists of materials that may be used to construct these types of props.

Other techniques are available and may be utilized in specialized applications. Some of these are briefly described in Section 12.

4. HEALTH STATES

Health states may be of temporary or chronic (i.e., permanent) duration. The preference for a health state is affected by its duration (Torrance et. al., 1972; Sackett and Torrance, 1978) and should not be confounded by preferences for other possible future states. To avoid these problems all states being compared must be specified as being of the same duration. States with different durations can be handled by using multiple batches. For a chronic condition the prognosis should be no change until death, with the age of death specified. For temporary states the prognosis should be described, in the task instructions, as no change until the end of the duration specified for the state at which time the person returns to perfect health (Torrance, 1982). Preferences for both chronic and temporary states may be measured using the rating scale, standard gamble, or time trade-off. The examples contained in this guide are limited to chronic states but the techniques are similar for measuring temporary states (Feeny et. al., 1990).
If it is important to measure preferences for temporary states they should be batched together to ensure common duration. Measurements would be made for each temporary state relative to perfect health and the worst temporary state of the group. If the programmes being evaluated involve only morbidity, not mortality, and there is no need to compare the results with programmes that involve mortality, the measurement process is complete. However, if all three conditions are not satisfied, the relative preference values for temporary states must be transformed onto the standard 0-1 dead-healthy utility scale. This is done by redefining the worst temporary state as a chronic state of the same duration and measuring its preference value relative to perfect health and death. The preference values for the other temporary health states are then converted to the standard 0-1 dead-healthy scale by using a positive linear transformation.

The basic method chosen to describe an individual health state to a respondent is determined by the type of study, the number of states, and the type of respondents(s) selected to evaluate the health states. If patient health states are the states to be measured and the patient is the evaluator, then the health-state description may be defined simply as the patient's subjectively-defined current health state (eg. "Your Current Health State"). However, if the study design requires the assessment of many different states or evaluation by respondents with no personal experience of the health state, it is necessary to describe hypothetical states that provide details of the impact of the health state on a person's ability to function, behave or perform.

There are many ways to describe health states, including narrative paragraphs, point form phrases and video tapes portraying people in health states of
interest. It is not the intent of this manual to provide a guideline for the development of health-state descriptions, but in general the descriptions should be easily understood by the intended respondent(s) and all important aspects of the health state must be included explicitly. Diagnosis, disease labels, laboratory test results and prognosis should not, in general, be included in health-state descriptions or as health-state descriptors. Cognitive burden increases directly with the number of attributes (i.e., components) used to describe each health state and is exacerbated when the number of states is large. Therefore it is suggested that the number of states to be evaluated at the same time be minimized. Fischer argues that a maximum of five attributes be used to describe each state (Fischer, 1979). Cognitive burden can also be reduced by using symbols or colour codes to highlight differences among states. Readers interested in the development of health-state descriptions are referred to Torrance (1982, 1986 and 1987).

The descriptions of health-state scenarios may be based explicitly on the multiattribute utility function approach (see Torrance et al., 1982). In this approach health states are defined by combinations of attribute functional levels. A set of basic attributes would consist of physical mobility, self-care, emotion, sensory, cognition, pain and other relevant characteristics (for example see Figure 1.2(a)). Each attribute is sub-divided into levels ranging from normal to severely limited. Combinations of levels among attributes provide for a wide range of health states. Even if the multiattribute approach is not used explicitly, it provides a useful implicit framework for constructing holistic health-state descriptions (for example see Figure 1.2(b)).
Studies that involve the evaluation of many different health states by each respondent will require a method to store and organize the descriptive cards or sticks. In most circumstances it is possible to minimize the number of descriptive cards by using the same cards for the feeling thermometer and standard gamble or time trade-off measures. Ideally the storage method will reduce unnecessary sorting and searching activities during the interview (i.e., will improve efficiency of interview time), will identify any missing descriptions before the interview commences and will protect the descriptions from damage during transit. Many storage methods are available and the most effective will depend upon the physical size and shape of the descriptions. Past studies have used plastic bags, cardboard folders and plastic parking permit holders. Each section (e.g., bag or holder) should be clearly labelled, reasonably easy to access and referred to at the appropriate point in the Schedule (see Appendix III for examples). Health states may be identified by codes to ease the recording of subject responses but the codes should be carefully selected to avoid influencing preference measures. Randomly selected letters, printed on the reverse side of health-state description cards are most often recommended.

5. INSTRUMENT TESTING

All interviewing materials should be thoroughly tested prior to the initiation of study data collection. Schedules and other written or verbal respondent instructions may be examined by a commercially available software package (e.g., RightWriter) to ensure that the level of readability is suitable for the target population. All components of the instrument should be included in the testing to ensure that the interview is not so lengthy that participants (i.e., interviewers and respondents) become tired, that all instructions are clear, and
that all required materials are available. In general, it is recommended that the average interview duration should not exceed 60 minutes. As a rule of thumb, the introduction and feeling thermometer ratings will take 20 to 30 minutes, and standard gamble (i.e., chance) or time trade-off questions will each require about 5 minutes. Therefore an interview that averages one hour may consist of a feeling thermometer section, 4 or 5 standard gamble questions, a few demographic questions, and a small number of interview evaluation questions.

Testing a prototype instrument is most efficient if study staff administer the draft interview to a small number of friends or colleagues and note any problems with the wording of the schedule, interviewer instructions, response marking forms or props. The instrument should then be revised and retested especially if the revisions were numerous, complex or involved major changes. It is often necessary for project staff to repeat the revision-retesting cycle many times before a complex instrument (eg. an instrument with many branching or skip questions) is ready to be formally pretested by the study interviewers.

Formal pretesting is a vital step in the development of any instrument because it is often the first occasion that the instrument is exposed to a critique of individuals (eg. interviewers and respondents) who have not been involved in its preliminary development. Interviewers must be briefed on the purpose of the study, trained to use the instrumentation and instructed to administer the instrument to a sample of respondents who are representative of the study population. Ideally two or three senior interviewers should each conduct a minimum of two interviews. After completing the four to six interviews the interviewers should be debriefed, if possible in a group setting, by project
staff to identify all outstanding problems encountered. The debriefing session should be structured to encourage the pretest interviewers to report all problems, even those of a seemingly trivial nature, and suggest improvements. Reports received during debriefing sessions provide much of the material contained in the Interviewer Manual (see example in Appendix II). In the event that major revisions to the instrument are required, it may be necessary to repeat formal pretesting one or more times until the interviewers are convinced that the instrument can be administered to the target population in an efficient manner and the researchers are satisfied that the study objectives will be met.

6. INTERVIEWERS

Selecting the person(s) who will administer the interview is an important aspect of any study (Fowler and Mangione, 1986). All data are collected by interviewers and therefore the quality of the data will be determined by the effectiveness of the interviewer. In general, the effectiveness of an interviewer will depend on her/his ability to:

(a) make a positive first impression on the respondent to encourage cooperation;

(b) develop effective communication with the respondent to ensure continued assistance;

(c) avoid influencing the respondent's opinion (i.e., avoid imparting any biases); and

(d) collect and record (clearly and accurately) the responses obtained during the interview.

It is recommended that interviewers for health-related studies be mature adults, experienced at obtaining and maintaining subject co-operation, and highly trained in the use of the utility instrument. (They need not, however, be familiar with the underlying conceptual foundations.) Training is required to ensure that all
Interviewers are familiar with the study objectives, instruments and any important target respondent characteristics. Training sessions should be conducted in small group settings and led by the interviewers' field supervisor.

Field supervisors should be thoroughly briefed in advance and would generally begin the training session by introducing the research group, who then explain the purpose and importance of the study. The field supervisor then provides each interviewer with a copy of the instrument and continues by summarizing the approach as outlined in the interviewer manual. Each interviewer is then asked to prepare their props and imagine themselves as a respondent while the field supervisor reads the schedule aloud and demonstrates the correct use of the props. At the end of the presentation, interviewers are asked to form pairs and alternate as respondent and interviewer to ensure that each interviewer completes at least one practice interview. This process may identify unresolved questions or problems and helps to clarify instructions. Professional interviewers will undertake additional practice interviews with family or friends to ensure that they are adequately prepared to administer the instrument to study respondents.

After the survey has been completed the interviewers are asked to attend a debriefing session. Debriefing provides the researchers with the opportunity to determine the overall quality of the survey, identify general problems that may be improved in any subsequent version of the instrument, and double-check missing or questionable recorded values.
7. Respondents

Because utility values are estimates of individual preferences for health states, the selection of respondents (i.e., subjects) will affect the potential generalizability of the results. There is evidence (Sackett and Torrance, 1978) that patient-utility values systematically differ, at least in some cases, from those reported by healthy members of the general public, although differences are usually not dramatic. Therefore the choice of respondents will most often depend upon the objectives of a study. For example, in a clinical decision-making situation it is likely that a patient utility value is the only relevant viewpoint, while in studies related to policies of publicly-funded health care services it may be necessary to include both patient and general public (i.e., societal) viewpoints. Drummond et al. (1987) suggest that patients and health professionals can often provide utility values based on personal experience, without the aid of detailed health-state descriptions, but are exposed to a potential conflict of societal interest because both these groups have an incentive to exaggerate the disutility of health states to enhance the cost-utility of their preventive and treatment programmes. However, patient or professional measurements are often of interest and including these measures often helps add credibility to results based on general-public subjects. Patient utilities may also be the most pertinent measures if all the alternatives being compared are directed at the same disorder, the utilities associated with the disorder are the only utilities required, and the results will not be compared with programmes directed at other disorders.

Ideally, patients are asked to provide measures for all relevant health-state descriptions, plus a measure of their own subjectively-defined current health
state. To assess chronic states, patients should be approached after treatment has been completed and their health status has stabilized. Ideally, patient preferences for temporary treatment process health states should be measured while the patient has the disease and is under-going treatment so that their subjectively-defined current state is represented. However, in practice it is difficult to approach patients or their families during the most stressful phases of a disease or treatment process. Therefore, it is often necessary to measure patients’ preferences after recovery and use health-state descriptions to help patients recall the details of their experiences. The number of respondents required to satisfy study objectives varies and is discussed in detail in Section 7.1 of this manual.

Some studies require the collation of many respondent variables and the follow-up of a population to determine compliance rates. In these relatively complex situations it is useful to create, using a word processing software package, a set of respondent lists as follows:

i) a master list of all potential respondents, ordered by survey identification number;

ii) a master list of all potential respondents in alphabetical order by surname;

iii) lists of randomly selected target respondents for each interviewer; and

iv) lists of replacement respondents, in random order, for each survey stratum (i.e., patient and general public, male and female etc.).

Replacement respondents are used when an excess number of potential subjects are available. In these circumstances, a target respondent who is unable or unwilling to participate in the survey would be replaced by a subject randomly selected from the same stratum as the original target respondent. Replacement lists are normally retained by the survey field co-ordinator and replacement
subjects are allocated to interviewers as required.

Ideally the lists should be created so that the information can be extracted in various formats. Respondent compliance is improved if an introductory letter is received by the target subject just prior to contact by the interviewer and this timing is best accomplished by allowing the interviewer to mail the letters according to their personal schedule. The letter should include a personalized salutation, an original signature of an investigator and be sent by first class mail. These, and other strategies for increasing consent rates are summarized by Woodward et. al. (1982).

7.1 Sample size, parameter estimates and hypothesis testing*

Although utility measures are reliable and valid, single measurements are not highly precise. For example, Torrance (1976) reports an estimate of the standard deviation of within-subject measures of 0.13. In addition, utility values for the same health state vary across individuals as indicated by estimates of standard deviations of 0.3 and 0.2 for general public and patient groups, respectively (Torrance, 1976). It is hypothesized that subjects, such as patients or parents of pediatric patients, who are experienced with specific health problems have a greater understanding of the impact and importance of health problems on quality of life than less experienced respondents.

*This section is lengthy and may be omitted by readers who do not require detailed knowledge on sample requirements.
Experience appears to result in more homogeneous responses, and associated reductions in standard deviation estimates, for groups of experienced subjects as compared to inexperienced subjects selected at random from a general-public population.

Fortunately, in the context of program and clinical evaluations it is group responses that are of most interest. Responses may be grouped by health state, or type of subject within health states. Because health-state utilities are measured quantities with interval scale properties, the best estimate of group response is the mean (Torrance et. al., 1982). Although there may be considerable differences among individuals within groups, the overall group mean estimate can be made as precise as required by increasing group size. To estimate the true value of a group mean the usual procedure is to select a random sample from the population of interest, calculate the mean response for the sample, and then calculate the confidence interval of the underlying population mean.

The confidence interval contains all values between two boundary points in which the true population mean is expected to lie, with a specified level of confidence. For example, a 95% confidence interval for a population mean would consist of lower and upper limits determined so that, in repeated sets of samples of the same size, 95 percent of all intervals would be expected to contain the population mean. The confidence interval for a population mean is given by the
The general formula:
\[ \bar{x} \pm [t_{\alpha/2, n-1}] \times \sqrt{\frac{S^2}{n}} \]
where \( n \) is the sample size,
\( \bar{x} \) is the sample mean,
\( t \) is the critical value on the t-distribution for \( n-1 \) degrees of freedom and a Type I error rate of \( \alpha \), and
\( S^2 \) is the sample variance.

Table 1 presents examples of approximate confidence intervals for population means of experienced (e.g., patients) and inexperienced (e.g., general public) subject group responses.

The confidence interval for the true difference between two population means is given by the formula:
\[ (\bar{x}_1 - \bar{x}_2) \pm [t_{\alpha/2, n_1 + n_2 - 2}] \times [S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}] \]
where \( \bar{x}_i \) is the mean of group \( i \);
\( n_i \) is the sample size of group \( i \);
\( t \) is the critical value on the t-distribution for \( n_1 + n_2 - 2 \) degrees of freedom and a Type I error rate of \( \alpha \), and
\[ S_p = \sqrt{\frac{S_1^2 + (n_2 - 1) S_2^2}{n_1 + n_2 - 2}} \]
\( S'_p \) is the pooled sample variance where
\[ S'_p = \frac{(n_1 - 1) S_1^2 + (n_2 - 1) S_2^2}{n_1 + n_2 - 2} \]

If the confidence interval for the true difference contains zero it implies that there is no significant difference between the population means at the specified (i.e., \( \alpha \)) level of confidence. If zero is not contained in the confidence interval for the true difference there is evidence to conclude that there is a significant difference. A confidence interval may be used to test other
hypotheses. Often, however, analysts prefer to restrict the amount of information required to interpret results and, therefore, may choose to use classical hypothesis tests (such as Student's t-test).

Provided group data fulfill basic assumptions, the appropriate (i.e., most efficient and powerful) statistical test for investigation of differences between two or among more than two population means is a t-test or analysis of variance respectively. The ability of a statistical test to detect a real difference among means, if it exists, is known as test power and is the complement of the probability of making a Type II error (i.e., concluding that there is no difference between treatments when in truth there is a real difference). Test power will depend on group sizes, whether the test is one or two-tailed, whether individual responses across groups are independent or correlated, the minimum difference that is considered to be meaningful (e.g. clinically important), and the Type I error rate (i.e., the probability of concluding a difference between groups exists when in truth there is no real difference). Unless there is a compelling reason to do otherwise, all tests of significant differences among means should be two-tailed, and with the exception of group size, the other factors that influence test power are under the control of the study investigator.

The minimum meaningful difference defines the minimum size of the real difference in means, between groups, that the analyst decides should be detectable by a test. In general, it will vary with the objectives of the study. A useful, but arbitrary rule of thumb in the context of utility measurement is to assume that 0.1 is the minimally important difference. In other words, differences of 0.1
or larger in utility scores between groups are meaningful.

The direct relationship between sample size and test power, all other factors being equal, makes it possible to calculate either the required sample size for a pre-specified test power or, to calculate test power for a given sample size. The latter calculation is useful when subjects such as patients are a scarce resource, while the former calculation can be used to determine the minimum sample size requirements in a less restricted population (e.g., general public).

Taylor (1983) suggests the following equations for sample size calculations.

I. Comparing Independent Group Means:

number per group = \( 2 \times (Z_a + Z) \frac{\sigma}{d} \).

II. Comparing Related Group Means:

number of pairs = \( (Z_a + Z) \frac{\sigma}{d} \);

and \( \sigma_d \) may be estimated as

\[ \sigma_d = \sqrt{\sigma_1^2 + \sigma_2^2 - 2 \rho \sigma_1 \sigma_2}. \]

Similarly, power may be estimated using the following equations.

III. Power of Tests Among Independent Groups with Homogeneity of Variance:

\[ Z_s = \frac{d}{\sigma \sqrt{(1/n_1 + 1/n_2)}} - Z_a. \]

IV. Power of Tests Among Independent Groups with Unequal Population Variances:

\[ Z_s = \frac{d - Z_a \sqrt{(\sigma_1^2/n_1 + \sigma_2^2/n_2)}}{\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}}. \]

V. Power of Tests Among Related Means:

\[ Z_s = \frac{d}{\sigma_d \sqrt{(1/n)}} - Z_a. \]

Where \( n \) = number of subjects (or pairs of subjects for paired designs);
\[ \sigma = \text{standard deviation among subjects in all groups;} \]
\[ \sigma_1 = \text{standard deviation among subjects in group 1;} \]
\[ \sigma_2 = \text{standard deviation among subjects in group 2;} \]
\[ \sigma_d = \text{standard deviation of differences between pairs of observations;} \]
\[ d = \text{minimum clinically important difference (absolute value);} \]
\[ z_\alpha = \text{standard normal deviate corresponding to a probability } \alpha \text{ of making a type 1 error;} \]
\[ z_\beta = \text{standard normal deviate corresponding to a probability } \beta \text{ of making a type 2 error.} \]

Tables 2 to 6 present examples of power calculations for different sample sizes of independent and correlated (e.g., paired) groups of experienced (e.g., patient) and inexperienced (e.g., general public) respondents. All calculations are based on a 5 percent Type I error rate (i.e., \( \alpha = 0.05 \)), a minimum important detectable utility score difference of 0.1, and two-tailed tests of significance. It is generally recommended that the minimum acceptable power be approximately 80 percent (i.e., maximum Type II error rate probability of 0.2) and therefore all examples were chosen to illustrate the sample size range around this criterion.

Table 2 presents test power estimates for balanced, independent groups of experienced respondents that may be expected to represent subjects who participate in clinical trials. To ensure that statistical tests have at least an 80 percent chance of detecting a 0.1 difference in mean utility scores it would be necessary to have at least 65 subjects in each treatment group. Table 3 presents estimates of statistical power for balanced, independent groups of general-public respondents who are likely to be recruited for convenience but, due to inexperience, are likely to present more variable responses. To ensure that statistical power is at least 80 percent it would be necessary to have 145 subjects per group to detect a difference of at least 0.1 in mean group utility scores.
Tables 4 and 5 present power calculations for paired responses from experienced and inexperienced subjects, respectively. Paired responses may be collected from subjects in cross-over studies or from subjects matched by a characteristic that would be expected to result in correlated scores (e.g. twins, parent-child groups). In cases where group responses cannot be assumed to be independent the statistical power, for a given sample size, will vary directly with the correlation among groups. To ensure 80 percent power to detect differences between paired groups of experienced subjects we would need approximately 55, 40, 30, and 15 pairs of responses for correlations of 0.2, 0.4, 0.6 and 0.8 respectively. For paired groups of general public subjects we would require approximately 120, 90, 60 and 30 pairs of responses to achieve adequate power for the same levels of correlation.

Table 6 presents calculated power values for comparisons of mean utility scores for independent, experienced vs. inexperienced respondent groups. This type of comparison is of interest when experienced subjects are scarce and the recruitment of inexperienced subjects is under consideration. In the event that, in the presence of adequate power, no significant difference between experienced and inexperienced group mean responses is detected, it may be reasonable to pool experienced and inexperienced responses. The pooled responses would increase the precision of mean health state utility estimates and increase hypothesis testing power for differences in mean utility of different hypothetical health states. From Table 6 it can be observed that if there are 30 or fewer patients available it will be practically impossible to recruit sufficient general public respondents to ensure adequate power to detect real differences. However, if 40, 50, 70, 100 or 150 experienced subjects are available, adequate hypothesis
testing power can be expected with 350, 200, 150, 125 and 95 general public respondents, respectively.

8. CONSENT
Generally consent is implicit when a respondent verbally agrees to participate in an interview. However, a signed written consent is often a condition of institutional ethics committee approval, especially for interviews involving patients who are potentially under duress. Most ethics committees require the use of consent forms that must be designed in accordance with regulations and be approved in advance of interviewing. In general it is advisable to ask all respondents to sign a consent form (see examples in Appendix I). The signed consent states that the respondent is a voluntary participant. This consent does not imply that the participant cannot withdraw, or that any names or personal data will be released. (All information received during interviews must remain anonymous and confidential.)

9. FEELING THERMOMETER RATING SCALE
The Feeling Thermometer, a type of visual analogue scale, is useful for assisting respondents to rank health states but does not directly provide valid cardinal utility measures (Parducci, 1974) and therefore should be considered a complement to standard gamble or time trade-off techniques (Keeney and Raiffa, 1976), not a stand-alone utility instrument. Ranking of health states is a required step preceding cascade-type standard gamble or time trade-off measurements. Ranking may also provide valuable consistency checks and improves measurement characteristics of the instrument (Laskey and Fischer, 1987).
9.1 DESIGN AND ASSEMBLY

This visual aid is called a Feeling Thermometer because it helps subjects to measure their feelings (i.e., preferences) for different health states. It consists of a simple interval scale from 0 to 100. Each scale value should be at least 5 mm from the adjacent values to permit narrow descriptive pointers (i.e., sticks, arrows etc.) to discriminate between values on the scale.

The scale, a high contrast photographic line print on resin-coated paper, is produced from artwork by a graphic art reproduction house. A non-slip (e.g., standard fabric store felt) background supports the pointers. (A complete listing of recommended materials is contained in Appendix VIII.) The background may be laid out on both sides of the scale or on one side only. The felt is provided to increase friction and prevent the pointers from accidentally sliding up or down the scale. Felt is preferable to velcro because the felt and scale are mounted on a stiff piece of reinforced foamboard or similar material. The top and bottom of the scale should be labelled with suitable anchor descriptions. In general it is recommended that the top of the scale (i.e. 100) be labelled "MOST DESIRABLE" and the bottom (i.e., 0) be labelled "LEAST DESIRABLE" (see Figure 1.3).

Pointers are often constructed of strips of balsa wood to be approximately 5 mm. deep (to allow them to be easily lifted off a surface), 5 mm. wide and 10 to 15 cm. long (to contain one line of typed text), depending upon the application. In a recent study, written descriptions of many health states were printed on separate cards (12.5 X 7.5 cm.) and these cards were colour coded and matched to a set of plastic-laminated paper arrows. The cards were formed by folding a pre-
cut page to form a pocket for the storage of the matching arrow (see Figure 1.2 (6)).

9.2 INTERVIEW SCHEDULES AND RESPONSE BOOKLETS

Appendix III presents an example of a combined feeling thermometer and chance board interview schedule. The schedule begins by providing the respondent with some general, background information about the research objectives and methods.

The feeling thermometer section of the schedule begins with the rationale for the name of the scale, followed by a general explanation of the characteristics of the scale and the introductions to health state descriptions. In general it is recommended that a description of perfect health be the first health state introduced to the respondent and the respondent be "forced" to assign perfect health to a scale value of 100. This exercise helps the respondent to develop a conceptual anchor of the most desirable limit of the scale. Ideally the respondent is then provided with all the remaining health-state descriptions, asked to identify the least preferable state, and assign this state to a scale value of zero. The least desirable scale limit has now been visually and conceptually anchored. In the event that there are too many states for the respondent to be expected to select a least desirable state, it may be necessary to pre-select a sub-set of states that may be expected (eg. from pretest results or expert opinion) to include the least preferable state. (In this situation subjects are allowed to change the order from that determined initially.) Additional sub-sets may also need to be defined to reduce respondent cognitive burden by introducing small groups of states simultaneously rather than one large set.
Appendix IV presents a response booklet that contains, as Table 1, a marking form for recording feeling thermometer scores for 14 health states. Each health state was identified by a code (eg. TG, SF, etc.). The codes for all health states were pre-printed to form the left-hand column and the right-hand column was reserved for the recording of feeling thermometer scores. In this example perfect health was assigned a score of 100 but the scores for all other states, including death, were left to the discretion of the respondent.

10.0 CHANCE BOARD

The chance board is used to elicit respondents' utility values based on the probability equivalence (Hershey et al., 1982) standard-gamble method.

The standard gamble is the classical method of measuring cardinal preferences under uncertainty, in the fields of economics and decision analysis, and is derived directly from basic axioms of utility theory as postulated by von Neumann and Morgenstern (1944, 1947). This method requires that the respondent be offered two alternatives. The two alternatives include three health-state outcomes which have been previously ranked by the respondent; the feeling thermometer is often a useful prop to elicit these rankings. Alternative 1 is described as uncertain with probability p of the most preferred health state and probability 1-p of the least preferred outcome. Alternative 2 is always described as a certain outcome (i.e., probability of 1) of the health state ranked, by the respondent, as intermediate relative to the states presented in Alternative 1. Although the theoretical basis for the standard-gamble method is well accepted in economic and decision theory, it has sometimes been considered to be difficult for respondents who have limited experience with the concepts of
The chance board was designed to communicate probabilities to respondents who have had little experience with the formal concepts. The board simultaneously presents the probabilities, of the two uncertain outcomes and one certain outcome, associated with the two options presented to respondents. Presentation of both uncertain outcomes may appear redundant at times but it reduces the potential risk of "framing" bias that may be associated with using only one state that may have a positive or negative connotation to the respondent (Read et al, 1984). The board uses diagrams of common gambling-type wheels with colour coded pie-shaped segments representing the probabilities. Colour matched health-state descriptions help respondents understand the options, and the probabilities have been transformed to per cent risk (i.e., chance). It is unusual for respondents to state an explicit indifference between the two choices, but indifference is a legitimate response and therefore the board has been designed to accommodate this possibility. The most common situation is that the respondent provides an implicit probability indifference value and it is important that the instrument respond in an effective manner to relieve the interviewer of the responsibility of identifying implicit indifference values or calculating utility scores.

To date this particular version of the board has been used successfully by approximately 15 interviewers to elicit responses from over 600 patient and general-public respondents involved in 6 separate health-care surveys conducted in southern Ontario, Canada.
10.1 Design and Assembly

The Chance Board consists of 3 main components as follows:

1) faceplate (Figure 2.1);
2) adjustable probabilities wheel (Figure 2.2); and
3) backplate (Figure 2.3).

Each of these components is constructed from bristol board. (A complete listing of recommended materials is contained in Appendix VIII.) The faceplate and chance wheel components must be printed, coloured and trimmed before laminating with plastic film. All alpha-numeric characters and arrows on the faceplate and chance wheel may be printed in black using a stencil or rub-off lettering. The only other colour required on the face plate is the green background of the Choice "B" chance value (i.e., 100) and Choice "B" spinner circle. The adjustable wheel requires the use of black alpha-numeric characters and borders plus pink and blue backgrounds to the chance value boxes and spinner circles.

The layout of the sequence of probability values and spinner circles, in a converging ping-pong (i.e., alternating back-and-forth between high and low values) strategy, was dictated by the need to reduce the possibility of (i) an anchoring bias whereby a strategy of constantly increasing/decreasing probabilities encourages respondents to "overshoot" (i.e., overestimate or underestimate) their indifference point, and (ii) a framing effect associated with consistently increasing or decreasing probabilities which are often interpreted by respondents as gains or losses relative to a point of reference. In some cases framing effects have been shown to cause reversals of preferences (Tversky and Kahneman, 1981). The green, pink and blue colours may be added with highlighter felt pens but a more professional effect is obtained with adhesive plastic films available from graphic art suppliers.
The codes along the periphery of the adjustable wheel are optional. The codes may provide a handy reference for interviewers although caution must be exercised in the interpretation of these codes in situations where the intermediate state is being assessed relative to states other than perfect health and death.

If perfect health and death are the descriptions that constitute choice "A", the codes will identify the estimated indifference probability, which is the utility value of the intermediate health state, at the "grip" cut-out (see item 3 on Figure 2.1). Three codes will appear in the "grip" cut-out at each chance option presented. If the respondent selects choice "A" option, the relevant code is that which appears closest to the choice "A" faceplate title (i.e., the uppermost code). If the respondent selects choice "B" option, the relevant code is that which appears closest to the choice "B" faceplate title (i.e., the lowest code). In the event that a respondent indicates indifference between choice "A" and "B", the relevant code would be the centre value. The code "C" signifies that the standard gamble question should be continued by rotating the adjustable wheel in a counter-clockwise direction, as indicated by the arrow above the grip cut-out (see item 6 of Figure 2.1), to solicit additional information leading up to identifying an indifference probability. Question mark codes represent responses that require additional clarification or inconsistent responses (see interviewer directions in example schedule).

If perfect health and death are not the uncertain outcomes presented in choice "A", the codes will provide the indifference probabilities for the state described in choice "B" relative to the outcome states presented in choice "A" but not relative to the standard dead-healthy (0-1) utility scale. However,
standard utility values may be computed if relative measures are made between all states and a common anchor state. The bottom anchor state is a health state that is expected to have a low utility score and is used in place of death. The anchor state is then measured relative to both perfect health and death to provide a reference point on the standard dead-healthy utility scale. This technique is commonly called "cascading" and is most often used in situations requiring measurements of numerous health states, in which it is viewed as important to avoid constantly confronting the respondent with death as an option, or in circumstances in which some of the health states have utility values near 0 or 1.0. There is evidence to suggest that the most accurate responses result from indifference values derived from the central 80 per cent of the scale (Kahneman and Tversky, 1982). For states that are expected to provide indifference values in the lower or upper 10% of the scale, the cascading approach may result in more accurate responses by transferring the measurement to a relative scale in which the measured indifference value would lie in the central 80 per cent of the scale (Farquhar, 1984). However, from a statistical perspective it is important to recognize that health state "cascade" utility measurements are not based on statistically independent events (Laskey and Fischer, 1987). The resulting parameter estimates will be unbiased but variance will increase with measurement serial order.

The health-state descriptions associated with the Chance Board alternatives should be concise, legible, and physically associated and colour co-ordinated with the probabilities presented on the board. Colour coding may be provided by using coloured paper, tinted overlay film, coloured stripes or coloured dots. Figure 2.1 presents a method that has been successfully used to temporarily
secure the health-state description cards to the board. If perfect health and
death always form the outcomes then simple descriptions and associated colouring
may be printed directly on the board. However, the prop is a more flexible tool
if plastic pockets (e.g., parking permit holders as indicated by item 5 on Figure
2.1) or velcro strips are used to allow a variety of health-state cards to be
used during "cascade" type measurements. Transparent plastic pockets may be
tinted with coloured, graphic art film to allow a single set of white coloured
health-state cards, used with the Feeling Thermometer, to "cascade" through the
set of pockets.

10.2 Interview Schedules and Response Booklets
Appendix III presents an example of a combined Feeling Thermometer and Chance
Board Interview Schedule. The chance board section of the schedule begins with
a general explanation of the board followed by a worked example. The worked
example describes a health problem and outcomes that are not related to the
interests of the study. (By using an unrelated example the respondent's opinions
about the study questions are not likely to be influenced.) The remainder of the
schedule provides the detailed verbal and non-verbal instructions to the
respondent plus directions to the interviewer for the relatively simple
evaluation of a health state, identified by an orange dot, relative to perfect
health and death. In this example no attempt is made to quantify the extent to
which the state may be considered worse than death. (States considered worse
than death are defined as having negative utility values on the conventional 0-1,
dead-healthy scale. Measurement of negative utility can be problematic (Torrance
1982, Torrance 1984, Torrance 1986. Further discussions about states considered
worse than death are contained on pages 37 through 39.) Note that although the
codes representing the indifference probabilities may be read directly from the circumference of the wheel they have also been included in the schedule for the convenience of the interviewer. A more detailed schedule is required to provide directions for the "cascading" of health states based on a respondent's own rankings, but the general concepts remain the same as the example provided.

The chance board response marking forms are included in the example response booklet presented as Appendix IV. The cover page identifies the respondent and basic interview characteristics. Table 1 provides space for recording 15 health-state ranking scores reported by a respondent during the feeling thermometer exercise. The chance board marking forms, Table 2, were designed to record the results from four standard-gamble questions in which the choice "B" states were less preferred than perfect health. As usual the state being evaluated, described as choice "B" (eg., orange dot card or yellow dot card or brown dot or red dot card), is measured relative to an uncertain outcome in choice "A" comprised of perfect health and death. Death must always be less preferred with respect to the choice "B" state, as indicated by the respondent during a ranking exercise (eg., feeling thermometer). Figure 2.4 presents the steps involved in a conventional standard gamble question. Figure 2.5 illustrates the position of health-state cards for evaluating intermediate states (i.e., SA, DQ, MU) during three "cascading" standard gamble questions. In this example the respondent would have previously ranked the health states in the following order: perfect health, SA, DQ, MU, death. Appendix V presents an example response marking form for three "cascading" standard-gamble questions.
10.3 Administration

Before beginning a standard-gamble question it is important to have determined the respondent's rank preferences, using a feeling thermometer or similar technique, for each health state to be measured. The health-state ranks are then used to ensure that the most preferable state is positioned on the left-hand side of the choice "A" section of the board, the least desirable state is located on the right-hand portion of choice "A" and the state of intermediate preference is associated with the choice "B" section of the board. For the purposes of this description it will be assumed that only three states are to be considered, perfect health will be the most preferable health-state, death the least preferable state, and the remaining state is less preferable than perfect health but preferable to death.

Once the health states have been arranged on the board the adjustable wheel should be rotated, using the crescent-shaped cut-out (Item 3 of Figure 2.1), to ensure that choice "A" is described as having a 100 per cent chance of perfect health and zero chance of death. The respondent is then asked to choose either choice "A" or choice "B" as indicated on the board. Obviously choice "A", with a 100 per cent chance of perfect health, is the only rational (i.e., consistent with rankings) choice in this first step. (Step 1 may appear unnecessary but it provides the respondent with a conceptual anchor of the maximum probability of perfect health, helps the respondent to feel comfortable with the task and allows the interviewer to evaluate the respondent's level of comprehension.) The characters appearing in the crescent-shaped cut-out, on the left-hand edge of the board, will provide directions to the interviewer as to how to proceed. If choice "A" has been selected the interviewer would rotate the adjustable wheel
in a counter-clockwise direction until the next complete set of probabilities appear in the cut-out sections of choice "A" as directed by the "C" (i.e., continue) in the upper portion of the crescent-shaped cut-out. Code "C" corresponds to the respondent having chosen the option presented in the upper portion of the board (i.e., choice "A"). Similarly, the code corresponding to a choice "B" response is represented by the characters in the lower section of the crescent-shaped cut-out and an indifferent response corresponds to the central code of the cut-out. For example, an indifferent response to the Step 1 prompt would be recorded as a utility score of 1.0 (i.e., there is no important difference in respondent preferences between the utility of perfect health and that described by the health state of choice "B") but a respondent who chose choice "B" would appear to have made an inconsistent decision which corresponds to the single question mark code.

Assuming that choice "A" was chosen in step 1, the interviewer would be directed to continue to step 2. Step 2 would be prepared by rotating the wheel in a counter-clockwise direction until choice "A" is described as a 10 per cent chance of perfect health and a 90 per cent chance of death. Again the respondent is prompted to chose either option "A" or "B" or to be indifferent between them. If choice "B" is chosen, the "C" code indicates the interviewer should continue to step 3 by rotating the wheel in a counter-clockwise direction. If respondent indifference is indicated the question ends and the score must be recorded as 0.1. In the event that choice "A" is chosen the corresponding code is "??", which indicates that the question ends after the respondent is prompted to report whether he/she considers choice "B" as more preferable, less preferable or equally preferable to certain death. If more preferable, the score is 0.05; if
equal, the score is 0.0; and if less preferable, the score is less than 0.0.

There are a maximum of 10 steps in each standard gamble question. The recommended pattern for the adjustable wheel will ensure that the probabilities of both perfect health and death will be presented in a converging "ping-pong" order provided that the interviewer continues to rotate the wheel counterclockwise with stops at each set of probabilities. Figure 2.4 presents a flow diagram of the potential steps and outcomes associated with a conventional standard gamble question that defines choice "A" as an uncertain outcome of perfect health or death.

11. TIME TRADE-OFF BOARD

The time trade-off method was developed specifically for use in health-care evaluations by Torrance et. al. (1972) and was devised to simplify health-state utility elicitation interviews. It is an implicit technique, like the standard gamble, that derives preference values implied by a subject's responses to decision situations (Torrance, 1986) and has been validated against the standard gamble method (Torrance et. al., 1973; Torrance, 1976).

Conceptually, for chronic states considered better than death, a respondent is asked to state the length of life in perfect health which he/she feels is equivalent to spending a longer period of time in a less desirable health state. The time trade-off scores are equivalent to utility (i.e., standard gamble) scores if the subject has a utility function that is linear in life years and if time is taken into account in standard gamble, and risk taken into account in time trade-off.
The Time Trade-Off Board is designed to convey the total years of life remaining and the years of life lost due to a premature death for each of two options. By presenting both these aspects the potential "life vs. death" framing bias, associated with the use of either the positive (years remaining) or the negative (years of life lost) aspect, is minimized. Another potential source of bias is associated with constantly increasing or decreasing the number of healthy years offered. This encourages the subject to "overshoot" his/her indifference point and results in an overestimate or underestimate of utility. This "anchor" bias is reduced by alternating between long and short durations of life using a converging ping-pong approach. For example, the respondent may initially be offered 40 years of additional life followed by 5, 35, 10, 30, 15, 25 and 20 years respectively. Of course, once the indifference point is identified all remaining potential durations are redundant.

11.1 Design and Assembly

This board may be separate from or integrated with (i.e., utilizing the reverse side of) a Chance board. A separate, or dedicated, time trade-off board consists of 4 separately constructed pieces as follows:

i) faceplate;

ii) backplate;

iii) "healthy" slider; and

iv) intermediate state slider.

Each piece of bristol board (cardboard) is printed, coloured and trimmed before laminating with plastic film (keeps the cardboard clean and resists damage). The plastic film is then trimmed to within one-eighth inch of the cardboard. The faceplate is adhered to the backplate with double-sided foam tape which forms the
slider guides, and secures the top and bottom edges (see Figures 3.1 and 3.2). Slider details are presented in Figure 3.3. (A complete listing of suggested materials is contained in Appendix VIII.)

In the case of our example in Figure 3.1, the upper section of the board is labelled LIFE A and the lower section is labelled LIFE B. LIFE A health state is preferred, as indicated by the respondent using a feeling thermometer or similar instrument, to the health state described in LIFE B. Both A and B choices are described in terms of health state, duration of life, time of death and period of life lost due to premature death. Health-state descriptions vary widely and the choice of content-format must be carefully considered. This board has been designed to accommodate 12 cm. X 8 cm. cards that are sufficiently large to accommodate 12 to 14 lines of text but limited to discourage the provision of overwhelming amounts of information. The health-state cards may be taped to the left-hand side of the board as indicated in LIFE A, or temporarily held in place by velcro strips, or placed in transparent plastic pockets as illustrated in LIFE B (see Figure 3.1).

The maximum period of life that may be traded is described by a horizontal scale to the right of the health state description cards. This maximum period is presented in terms of life expectancy, determined from age and gender specific life tables. Life expectancy may be described in terms of a relative scale of expected remaining time or as an absolute scale presenting respondent ages. The scale and its intervals must be selected to maximize measurement efficiency given the population of interest and study objectives. For example, in a pediatric population with an average age of four years and an actuarial life expectancy of
an additional 66 years, an absolute scale would have a minimum value of 4 years
of age and maximum value of 70 years of age. An analogous relative scale would
describe expected years of life remaining and have minimum value of 0 years and
a maximum value of 66 years. The absolute scale has the advantage of more
clearly presenting the respondent's current age, conveying the hypothetical age
of death and the period of life lost due to premature death. However, the
relative scale of remaining years is more flexible for use in studies that
involve respondents whose ages vary and the respondents' indifference values may
be identified more directly. The scales may be printed directly onto the bristol
board and/or temporarily affixed with velcro tapes, double-sided tape or other
similar devices although care is required to avoid board damage and poor
aesthetics associated with removable scales.

Figure 3.4 provides some examples of scales. The two stage scale would employ
a series of converging ping-pong questions using the upper 5-unit interval scale
to identify efficiently the specific 5-unit interval for second stage questions
using the 1-unit intervals. Scales 10 and 11 may be useful to make more precise
measurements in a predefined section of the scale expected to be most relevant.

Periods of perfect health, less desirable health and life lost due to premature
death are colour coded to ease respondent cognitive burden. The colours chosen
should be easily differentiated and selected to be meaningful to target
respondents. For example, Figure 3.3 sliders depict periods of perfect health
by a pink colour as related to the phrase "in the pink of health", less desirable
health by a light blue colour associated with "feeling blue", and life lost to
death by black as is traditional in many Western cultures. The asterisks on
sliders A and B are used to indicate the time of death. For most applications slider B would remain fixed such that the asterisk remains at the maximum scale value. However, the dot on slider B provides an interviewer reference for presenting the transition between perfect health and less desirable health in situations requiring the measurement of states considered by the respondent to be worse than death (Torrance 1984 and Cadman et. al 1986).

11.2 Interview Schedules and Response Booklets

Appendix VI presents an example of a time trade-off section for an interview schedule and Appendix VII presents a recording form from a response booklet. These materials are based on a hypothetical study involving respondents with six years of additional life expectancy (eg. a geriatric population). The schedule provides the interviewer with a verbatim text to guide the respondent through the time trade-off question and supplementary directions for the interviewer to provide the respondent with non-verbal prompts.

The response marking form has been designed to record the respondent’s answers to all choices presented and to provide a quick-reference guide to the board format at each step. Interviewers who are familiar with the time trade-off process may feel that recording responses for each choice is unnecessary but most professional interviewers report that the structured recording form provides a valuable double-check during a potentially confusing series of steps. The table is to be marked by the interviewer. The table consists of two columns of years, or ages if an absolute scale is used. The table rows show the two choices available at any one time. Responses are marked by placing a check mark between the brackets corresponding to the "Life" option selected by the respondent.
Normally one mark will appear on each row and subsequent "Life" choices may be determined by following the arrow originating from the marked response to the row or page indicated. In the event that a respondent considers the "Life" choices equal, a check mark is placed beside both options and the interviewer proceeds to the next question. The "/50-52", at the bottom right-hand corner of the response marking form is for office use only and is a data management code that specifies the columns of the database array assigned to the utility value calculated for the time trade-off question.

11.3 Administration

Each time trade-off question presents the respondent with a series of two options. The options are labelled Life A and Life B. Life A is generally described as perfect health, the duration of life in the perfect health state, the time of death and the duration of life lost due to premature death. Life B is described by a less preferred health state that would typically last the maximum duration under consideration. However, if the less preferred state described in life B is considered to be worse than being dead and the extent to which the state is considered worse than death is of interest, the board will allow for presentation of a duration of full health followed by a duration of the less preferred state of health. It should be noted that estimates of negative utility values (i.e., for states worse than death) are problematic and great care should be taken in the interpretation of these results (Torrance 1982, Torrance 1984, Torrance 1986). (A technique commonly used to avoid the need to interpret negative utility values is to define death, equal to zero, as the lower bound of the utility scale. Therefore negative values are undefined and constrained by
the zero boundary to equal zero.)

The order in which the series of options are presented is important. A converging "ping-pong" approach is recommended to eliminate the anchoring and framing biases associated with the presentation of minimum or maximum trade-offs. The ping-pong series begins with zero healthy time traded, followed by a maximum trade-off, followed by zero plus one unit trade-off, followed by maximum minus one unit trade-off and continues this pattern until the respondent's indifference trade-off value is identified. The indifference value is reached when the respondent states that she/he is indifferent between two choices or, when the respondent rejects life A of duration X units but accepts life A of X plus one unit. (It is unusual for respondents to explicitly state an indifference between choices and therefore it is very important to be prepared to identify the implicit indifference point.)

A general example of the technique is presented as Figure 3.5. A simple relative scale, varying from 0 to 6 years, is used to illustrate the series of options presented. The calculated indifference values assume that the less preferred state is considered equal to or better than death, or that death equals zero and zero has been defined as the lower bound on the utility scale (i.e., the negative scale is presumed to be undefined). The units of time are defined as additional years and the interviewer begins by presenting the respondent with a choice of 6 years of life A or 6 years of life B. Because life B health state has been defined by the respondent as less preferable than life A, it would be an inconsistent response if the respondent was indifferent between life A and life B or if the respondent chose 6 years in life B rather than 6 years in life A.
Assuming that life A was chosen, the interviewer would proceed to step 2 by sliding the asterisk of life A slider to zero years (i.e., immediate death) and eliciting the respondents choice. If the respondent chose life A, rather than 6 years of life B, this implies that life B is worse than dying immediately. The utility of the health state for life B would be zero, because the negative scale is undefined, and the series of choices would end. However, if 6 years of life B was selected in Step 2, rather than zero years of life A, the interviewer would continue to Step 3 by moving the asterisk of the life A slider to indicate death after 5 years of perfect health. If life B was chosen, the respondent is rejecting the 5 years of perfect health and one year of life lost due to premature death but accepting 6 years of perfect health in step 1. Therefore, the indifference value would lie between 5 and 6 years and be estimated as the mid-point value of 5.5 years. In this simple example there is the potential requirement for a series of 7 separate choices. The sequence, formats and results for all potential steps in this example are presented in Figure 3.5.

The preceding example is considered a simple example for the following reasons:

1) the health state of life A is perfect health which is defined as the upper bound of the conventional dead-healthy scale (utility 0.0 through 1.0);

2) negative utility values are considered undefined and therefore slider B doesn’t move and is fixed at death after 6 years;

3) the number of intervals in the scale is small (i.e., six); and

4) a relative scale is used.

The time trade-off approach and board may also be used when these four conditions are not satisfied. The health state of life A must be preferred to the health state of life B but need not be defined as perfect health. If an intermediate state, I, was substituted for perfect health then the calculated utility value
of the life B state, L, will be a measure relative to death and the state I, rather than being a measure on the conventional dead - healthy scale. The measure can be converted to the conventional scale by adding a second time trade-off question which would present life A as perfect health and life B as that used in life A from the previous question (i.e., state I). This technique is often referred to as chaining or "cascading" states through a series of time trade-off questions. The first time trade-off question would provide a preference measure for state L relative to death and state I, while the second time trade-off question would provide a preference measure for state I relative to perfect health and death. This "cascading" approach can, in theory, be extended to the evaluation of any number of health states but is limited in practice, by respondent burden, to approximately 6 states. In general it is recommended that life A be described by perfect health, to eliminate the serial dependence of values based on converted measures (refer to discussion on page 29), but this guideline is subject to exceptions.

12. OTHER VISUAL AIDS

Other prop formats, including a card deck and chart-like masking sheets, have been utilized successfully in special applications (Bombardier et. al., 1986 and Pauker et. al., 1988). These types of props are often simpler and less expensive to construct than boards. In addition, some interviewers consider it easier, especially in bedside settings, to flip through a set of cards rather than slide sections of a board. Figure 4.0 presents a diagram of a standard gamble question displayed on a card, contained within a set of cards that each present systematically a different question. (Sets of cards may also be used to display time trade-off information.) Typically the set of cards is secured by a ring or
similar device, to ensure that the questions are presented to every respondent in the correct order. Codes may be provided to describe the probabilities for the uncertain outcome alternative in the question (see upper right-hand corner of example card), and to identify final scores or continuation prompts (see codes across lower section of card) on the basis of responses by subjects.

Figure 4.1 provides a diagram of time trade-off questions presented in a fixed order using a chart and mask visual aid. This prop requires two sheets of paper (or card). The top (i.e., overlay) sheet consists of a title and a time scale, separated by a cut-out section. The lower sheet consists of a series of chart-like graphics that are displayed, one at a time, through the cut-out section of the top sheet. The chart-like graphics are drawn such that sequential questions will converge inward, in a "ping-pong" fashion, on the subject's indifference value as the lower sheet is drawn out from under the top (i.e., mask) page.

Recently, there have been attempts to program computers to present visual aid prompts, identify indifference scores and store responses in disc files. This approach eliminates manual recording and subsequent transcription of data but the required investment in hardware and software is high, the work must be customized for each application, and the approach is not yet readily accepted by some subjects and interviewers. In addition, computers cannot be expected to replace the skills of a professional interviewer. Interviewer leadership is generally required to convince subjects to participate, help respondents to feel at ease, clarify tasks and health state descriptions, encourage thoughtfulness, administer instruments at the subject's convenience (eg., time and place), and evaluate the quality of responses. It is anticipated that in the future computers will be
valuable tools for the collection of utility data but not until suitable software is available and computers are less intimidating to the majority of subjects and interviewers.

13. CONCLUSIONS

In sum, with careful instrument development and pre-testing, reliable, valid, and responsive utility scores may be obtained for health states. These scores can be useful in the evaluation of outcome in terms of health-related quality of life. Utility scores are also useful in evaluations such as cost-utility analyses in which the costs of health care interventions are compared to their consequences both in terms of the quantity and quality of life. In order to obtain useful scores, however, utility elicitation interviews need to be carefully planned and executed. Experience to date indicates that this can be done but also that it requires time, effort and care.
BIBLIOGRAPHY


Torrance GW. Health States Worse than Death. In: Eimeren WV, Engelbrecht R and Flagle ChD, eds. 3rd Int. Conf on System Science in Health Care. Berlin: Springer Verlag, 1984; 1085-1089.


### TABLE 1

**Examples of approximate 95% confidence intervals for population means**

<table>
<thead>
<tr>
<th>EXPERIENCED SUBJECTS</th>
<th>S.D. = 0.2</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>n</td>
<td>t-value</td>
<td>interval $^1$</td>
</tr>
<tr>
<td>----</td>
<td>-------</td>
<td>-------------</td>
</tr>
<tr>
<td>10</td>
<td>2.26</td>
<td>0.1429</td>
</tr>
<tr>
<td>20</td>
<td>2.10</td>
<td>0.0939</td>
</tr>
<tr>
<td>30</td>
<td>2.05</td>
<td>0.0749</td>
</tr>
<tr>
<td>40</td>
<td>2.02</td>
<td>0.0639</td>
</tr>
<tr>
<td>50</td>
<td>2.01</td>
<td>0.0569</td>
</tr>
<tr>
<td>60</td>
<td>2.00</td>
<td>0.0516</td>
</tr>
<tr>
<td>70</td>
<td>1.96</td>
<td>0.0469</td>
</tr>
<tr>
<td>80</td>
<td>1.96</td>
<td>0.0438</td>
</tr>
<tr>
<td>90</td>
<td>1.96</td>
<td>0.0413</td>
</tr>
<tr>
<td>100</td>
<td>1.96</td>
<td>0.0392</td>
</tr>
<tr>
<td>110</td>
<td>1.96</td>
<td>0.0374</td>
</tr>
<tr>
<td>120</td>
<td>1.96</td>
<td>0.0358</td>
</tr>
<tr>
<td>130</td>
<td>1.96</td>
<td>0.0344</td>
</tr>
<tr>
<td>140</td>
<td>1.96</td>
<td>0.0331</td>
</tr>
<tr>
<td>150</td>
<td>1.96</td>
<td>0.0320</td>
</tr>
<tr>
<td>160</td>
<td>1.96</td>
<td>0.0310</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>INEXPERIENCED SUBJECTS</th>
<th>S.D. = 0.3</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>n</td>
<td>t-value</td>
<td>interval $^1$</td>
</tr>
<tr>
<td>----</td>
<td>-------</td>
<td>-------------</td>
</tr>
<tr>
<td>40</td>
<td>2.02</td>
<td>0.0958</td>
</tr>
<tr>
<td>60</td>
<td>2.00</td>
<td>0.0775</td>
</tr>
<tr>
<td>80</td>
<td>1.96</td>
<td>0.0657</td>
</tr>
<tr>
<td>100</td>
<td>1.96</td>
<td>0.0588</td>
</tr>
<tr>
<td>125</td>
<td>1.96</td>
<td>0.0526</td>
</tr>
<tr>
<td>150</td>
<td>1.96</td>
<td>0.0480</td>
</tr>
<tr>
<td>175</td>
<td>1.96</td>
<td>0.0444</td>
</tr>
<tr>
<td>200</td>
<td>1.96</td>
<td>0.0416</td>
</tr>
<tr>
<td>225</td>
<td>1.96</td>
<td>0.0392</td>
</tr>
<tr>
<td>250</td>
<td>1.96</td>
<td>0.0372</td>
</tr>
<tr>
<td>275</td>
<td>1.96</td>
<td>0.0355</td>
</tr>
<tr>
<td>300</td>
<td>1.96</td>
<td>0.0339</td>
</tr>
<tr>
<td>325</td>
<td>1.96</td>
<td>0.0326</td>
</tr>
<tr>
<td>350</td>
<td>1.96</td>
<td>0.0314</td>
</tr>
<tr>
<td>375</td>
<td>1.96</td>
<td>0.0304</td>
</tr>
<tr>
<td>400</td>
<td>1.96</td>
<td>0.0294</td>
</tr>
</tbody>
</table>

**Notes:**
1) "interval" value is added to, and subtracted from the mean to calculate upper and lower confidence interval bounds.
2) Recall that utility scores lie in the 0.0 to 1.0 interval.
TABLE 2

POWER CALCULATION DATA FOR COMPARISONS OF MEAN UTILITY SCORES FOR INDEPENDENT EXPERIENCED (eg., patient) RESPONDENT GROUPS (eg. males vs females).

FOR $d = 0.1$, $\alpha = 0.05$, TWO-TAILED TEST

<table>
<thead>
<tr>
<th>GROUP I (SD=0.2)</th>
<th>GROUP II (SD=0.2)</th>
<th>$Z_B$</th>
<th>POWER</th>
</tr>
</thead>
<tbody>
<tr>
<td>$n$</td>
<td>$n$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>100</td>
<td>1.575534</td>
<td>$&gt;0.939$</td>
</tr>
<tr>
<td>80</td>
<td>80</td>
<td>1.202278</td>
<td>$&gt;0.885$</td>
</tr>
<tr>
<td>70</td>
<td>70</td>
<td>0.998040</td>
<td>$&gt;0.828$</td>
</tr>
<tr>
<td>65</td>
<td>65</td>
<td>0.890439</td>
<td>$&gt;0.802$</td>
</tr>
<tr>
<td>60</td>
<td>60</td>
<td>0.778613</td>
<td>$&gt;0.773$</td>
</tr>
<tr>
<td>50</td>
<td>50</td>
<td>0.540000</td>
<td>$&gt;0.708$</td>
</tr>
<tr>
<td>40</td>
<td>40</td>
<td>0.276068</td>
<td>$&gt;0.599$</td>
</tr>
</tbody>
</table>
TABLE 3
POWER CALCULATION DATA FOR COMPARISONS OF MEAN UTILITY SCORES
FOR INDEPENDENT GENERAL PUBLIC RESPONDENT GROUPS
(eg. males vs females).

FOR $d = 0.1$, $\alpha = 0.05$, TWO-TAILED TEST

<table>
<thead>
<tr>
<th>GROUP I (SD=0.3)</th>
<th>GROUP II (SD=0.3)</th>
<th>$Z_\beta$</th>
<th>POWER</th>
</tr>
</thead>
<tbody>
<tr>
<td>n</td>
<td>n</td>
<td></td>
<td></td>
</tr>
<tr>
<td>300</td>
<td>300</td>
<td>2.122483</td>
<td>&gt;0.982</td>
</tr>
<tr>
<td>250</td>
<td>250</td>
<td>1.766780</td>
<td>&gt;0.959</td>
</tr>
<tr>
<td>200</td>
<td>200</td>
<td>1.373333</td>
<td>&gt;0.911</td>
</tr>
<tr>
<td>150</td>
<td>150</td>
<td>0.926751</td>
<td>&gt;0.816</td>
</tr>
<tr>
<td>145</td>
<td>145</td>
<td>0.878231</td>
<td>&gt;0.802</td>
</tr>
<tr>
<td>140</td>
<td>140</td>
<td>0.828867</td>
<td>&gt;0.788</td>
</tr>
<tr>
<td>130</td>
<td>130</td>
<td>0.727419</td>
<td>&gt;0.758</td>
</tr>
<tr>
<td>120</td>
<td>120</td>
<td>0.621989</td>
<td>&gt;0.725</td>
</tr>
<tr>
<td>110</td>
<td>110</td>
<td>0.512066</td>
<td>&gt;0.691</td>
</tr>
<tr>
<td>100</td>
<td>100</td>
<td>0.397023</td>
<td>&gt;0.636</td>
</tr>
</tbody>
</table>
TABLE 4
POWER CALCULATION DATA FOR COMPARISONS OF MEAN UTILITY SCORES FOR PAIRED EXPERIENCED (eg. patient) RESPONSES (SD = 0.2)

FOR $d = 0.1$, $\alpha = 0.05$, TWO-TAILED TEST

<table>
<thead>
<tr>
<th>$r$</th>
<th>$n$</th>
<th>SD diff</th>
<th>$Z_\beta$</th>
<th>POWER</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.8</td>
<td>25</td>
<td>0.1265</td>
<td>1.9928</td>
<td>&gt;0.974</td>
</tr>
<tr>
<td>0.8</td>
<td>20</td>
<td>0.1265</td>
<td>0.0500</td>
<td>&gt;0.939</td>
</tr>
<tr>
<td>0.8</td>
<td>15</td>
<td>0.1265</td>
<td>1.1019</td>
<td>&gt;0.864</td>
</tr>
<tr>
<td>0.8</td>
<td>10</td>
<td>0.1265</td>
<td>0.5400</td>
<td>&gt;0.691</td>
</tr>
<tr>
<td>0.6</td>
<td>35</td>
<td>0.1789</td>
<td>1.3472</td>
<td>&gt;0.903</td>
</tr>
<tr>
<td>0.6</td>
<td>30</td>
<td>0.1789</td>
<td>1.1019</td>
<td>&gt;0.864</td>
</tr>
<tr>
<td>0.6</td>
<td>25</td>
<td>0.1789</td>
<td>0.8351</td>
<td>&gt;0.788</td>
</tr>
<tr>
<td>0.6</td>
<td>20</td>
<td>0.1789</td>
<td>0.5400</td>
<td>&gt;0.692</td>
</tr>
<tr>
<td>0.4</td>
<td>50</td>
<td>0.2191</td>
<td>1.2675</td>
<td>&gt;0.894</td>
</tr>
<tr>
<td>0.4</td>
<td>45</td>
<td>0.2191</td>
<td>1.1019</td>
<td>&gt;0.864</td>
</tr>
<tr>
<td>0.4</td>
<td>40</td>
<td>0.2191</td>
<td>0.9268</td>
<td>&gt;0.816</td>
</tr>
<tr>
<td>0.4</td>
<td>35</td>
<td>0.2191</td>
<td>0.7403</td>
<td>&gt;0.758</td>
</tr>
<tr>
<td>0.2</td>
<td>60</td>
<td>0.2530</td>
<td>1.1019</td>
<td>&gt;0.864</td>
</tr>
<tr>
<td>0.2</td>
<td>55</td>
<td>0.2530</td>
<td>0.9715</td>
<td>&gt;0.829</td>
</tr>
<tr>
<td>0.2</td>
<td>50</td>
<td>0.2530</td>
<td>0.8351</td>
<td>&gt;0.788</td>
</tr>
<tr>
<td>0.2</td>
<td>45</td>
<td>0.2530</td>
<td>0.6917</td>
<td>&gt;0.742</td>
</tr>
</tbody>
</table>
TABLE 5
POWER CALCULATION DATA FOR COMPARISONS OF MEAN UTILITY SCORES FOR PAIRED GEN. PUBLIC RESPONDENTS \((SD = 0.3)\)
FOR \(d = 0.1, \alpha = 0.05\), TWO-TAILED TEST

<table>
<thead>
<tr>
<th>(r)</th>
<th>(n)</th>
<th>SD diff</th>
<th>(Z_B)</th>
<th>POWER</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.8</td>
<td>35</td>
<td>0.1897</td>
<td>1.1580</td>
<td>&gt;0.875</td>
</tr>
<tr>
<td>0.8</td>
<td>30</td>
<td>0.1897</td>
<td>0.9268</td>
<td>&gt;0.816</td>
</tr>
<tr>
<td>0.8</td>
<td>25</td>
<td>0.1897</td>
<td>0.6752</td>
<td>&gt;0.742</td>
</tr>
<tr>
<td>0.8</td>
<td>20</td>
<td>0.1897</td>
<td>0.3970</td>
<td>&gt;0.637</td>
</tr>
<tr>
<td>0.6</td>
<td>70</td>
<td>0.2683</td>
<td>1.1580</td>
<td>&gt;0.875</td>
</tr>
<tr>
<td>0.6</td>
<td>60</td>
<td>0.2683</td>
<td>0.9268</td>
<td>&gt;0.816</td>
</tr>
<tr>
<td>0.6</td>
<td>50</td>
<td>0.2683</td>
<td>0.6752</td>
<td>&gt;0.742</td>
</tr>
<tr>
<td>0.6</td>
<td>40</td>
<td>0.2683</td>
<td>0.3970</td>
<td>&gt;0.637</td>
</tr>
<tr>
<td>0.4</td>
<td>100</td>
<td>0.3286</td>
<td>1.0829</td>
<td>&gt;0.853</td>
</tr>
<tr>
<td>0.4</td>
<td>90</td>
<td>0.3286</td>
<td>0.9268</td>
<td>&gt;0.816</td>
</tr>
<tr>
<td>0.4</td>
<td>80</td>
<td>0.3286</td>
<td>0.7617</td>
<td>&gt;0.773</td>
</tr>
<tr>
<td>0.4</td>
<td>70</td>
<td>0.3286</td>
<td>0.5859</td>
<td>&gt;0.709</td>
</tr>
<tr>
<td>0.2</td>
<td>130</td>
<td>0.3795</td>
<td>1.0446</td>
<td>&gt;0.841</td>
</tr>
<tr>
<td>0.2</td>
<td>120</td>
<td>0.3795</td>
<td>0.9268</td>
<td>&gt;0.816</td>
</tr>
<tr>
<td>0.2</td>
<td>110</td>
<td>0.3795</td>
<td>0.8039</td>
<td>&gt;0.788</td>
</tr>
<tr>
<td>0.2</td>
<td>100</td>
<td>0.3795</td>
<td>0.6752</td>
<td>&gt;0.742</td>
</tr>
</tbody>
</table>
TABLE 6
POWER CALCULATION DATA FOR COMPARISONS OF MEAN
UTILITY SCORES FOR INDEPENDENT RESPONDENT GROUPS
FOR δ = 0.1, α = 0.05, TWO-TAILED TEST

<table>
<thead>
<tr>
<th>PATIENTS (SD=0.2)</th>
<th>GENERAL PUBLIC (SD=0.3)</th>
<th>n</th>
<th>n</th>
<th>S.E.diff</th>
<th>Z_0</th>
<th>POWER</th>
</tr>
</thead>
<tbody>
<tr>
<td>150</td>
<td>100</td>
<td>150</td>
<td>100</td>
<td>0.034157</td>
<td>0.967700</td>
<td>&gt;0.828</td>
</tr>
<tr>
<td>150</td>
<td>95</td>
<td>150</td>
<td>95</td>
<td>0.034843</td>
<td>0.910016</td>
<td>&gt;0.816</td>
</tr>
<tr>
<td>150</td>
<td>90</td>
<td>150</td>
<td>90</td>
<td>0.035590</td>
<td>0.849757</td>
<td>&gt;0.788</td>
</tr>
<tr>
<td>100</td>
<td>150</td>
<td>100</td>
<td>150</td>
<td>0.031623</td>
<td>1.202278</td>
<td>&gt;0.885</td>
</tr>
<tr>
<td>100</td>
<td>125</td>
<td>100</td>
<td>125</td>
<td>0.033466</td>
<td>1.028072</td>
<td>&gt;0.841</td>
</tr>
<tr>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>0.036056</td>
<td>0.813501</td>
<td>&gt;0.788</td>
</tr>
<tr>
<td>70</td>
<td>200</td>
<td>70</td>
<td>200</td>
<td>0.031960</td>
<td>1.168931</td>
<td>&gt;0.875</td>
</tr>
<tr>
<td>70</td>
<td>150</td>
<td>70</td>
<td>150</td>
<td>0.034226</td>
<td>0.961744</td>
<td>&gt;0.828</td>
</tr>
<tr>
<td>70</td>
<td>125</td>
<td>70</td>
<td>125</td>
<td>0.035936</td>
<td>0.822690</td>
<td>&gt;0.788</td>
</tr>
<tr>
<td>50</td>
<td>250</td>
<td>50</td>
<td>250</td>
<td>0.034059</td>
<td>0.976101</td>
<td>&gt;0.829</td>
</tr>
<tr>
<td>50</td>
<td>200</td>
<td>50</td>
<td>200</td>
<td>0.035355</td>
<td>0.868427</td>
<td>&gt;0.802</td>
</tr>
<tr>
<td>50</td>
<td>175</td>
<td>50</td>
<td>175</td>
<td>0.036253</td>
<td>0.798386</td>
<td>&gt;0.773</td>
</tr>
<tr>
<td>40</td>
<td>410</td>
<td>40</td>
<td>410</td>
<td>0.034922</td>
<td>0.903564</td>
<td>&gt;0.816</td>
</tr>
<tr>
<td>40</td>
<td>350</td>
<td>40</td>
<td>350</td>
<td>0.035456</td>
<td>0.860380</td>
<td>&gt;0.802</td>
</tr>
<tr>
<td>40</td>
<td>300</td>
<td>40</td>
<td>300</td>
<td>0.036056</td>
<td>0.813501</td>
<td>&gt;0.788</td>
</tr>
<tr>
<td>30</td>
<td>1000000</td>
<td>30</td>
<td>1000000</td>
<td>0.036516</td>
<td>0.778520</td>
<td>&gt;0.773</td>
</tr>
<tr>
<td>30</td>
<td>10000</td>
<td>30</td>
<td>10000</td>
<td>0.036638</td>
<td>0.769416</td>
<td>&gt;0.773</td>
</tr>
<tr>
<td>30</td>
<td>1000</td>
<td>30</td>
<td>1000</td>
<td>0.037727</td>
<td>0.690615</td>
<td>&gt;0.742</td>
</tr>
<tr>
<td>30</td>
<td>100</td>
<td>30</td>
<td>100</td>
<td>0.047258</td>
<td>0.156037</td>
<td>&gt;0.559</td>
</tr>
</tbody>
</table>
FIGURES
FLOW DIAGRAM OF BASIC STEPS IN DEVELOPMENT OF

HEALTH-STATE UTILITY MEASUREMENTS

PROTOCOL
(question, rationale, design, analysis, ethics, funding etc.)

WHOSE PREFERENCES?

WHAT PREFERENCES?

HOW TO MEASURE?

RESPONDENTS
- patients or proxy
- general public
- others

HEALTH STATES
- process/outcome
- number
- content
- comprehension level
- medium
- style
- format

SAMPLE SIZE
- important difference
- population variance
- test power
- level of significance
- between group correlation

SURVEY

TECHNIQUE
1) visual analogue scale, plus
2) standard gamble or time trade-off

PROPS
1) Feeling Thermometer
2) Chance Board/
   Time Trade-Off
   Board or
   PUHS Deck

DOCUMENTATION
1) schedule
2) recording forms
3) manual

PRETESTING
- to assess tactics
- by interviewers (2)
- representative sample (6)
- debriefing

PROTOYPE TESTING
- to assess strategy
- by study staff
- sample of convenience (approx. 4)
FEELING THERMOMETER BOARD SPECIFICATIONS

Scale 1:4
Able to see, hear and speak normally for age.
Requires mechanical equipment (such as canes, crutches, braces or wheelchair) to walk or get around independently.
Generally happy and free from worry.
Learns and remembers school work normally for age.
Eats, bathes, dresses and uses the toilet normally for age.
Free of pain and discomfort.
Able to have children with a fertile spouse.
Figure 1.2 (b)

LAYOUT OF A DESCRIPTIVE CARD FOR A SHORT DURATION

HEALTH STATE ASSOCIATED WITH A SPECIFIC DISEASE AND TREATMENT

1. You will live in hospital for one month and receive drugs, intravenous feeding and blood transfusions. On a few occasions you will need to visit another hospital and lie frustratingly still for radiation treatments or x-rays.

2. In the first week an operation will be performed to insert a plastic tube through the skin of your chest and along a vein to your heart.

3. During the fourth week you will have another major operation to remove tumours from your abdomen and chest.

4. On one occasion during this month a needle will be inserted into your pelvic bone to obtain a bone marrow sample.

5. You will be unable to move your legs and the doctors are not sure you will ever have the use of them again. For three weeks of this month you will have to remain in bed and be helped to wash, dress and use the toilet. During the third week you will be able to play for short periods but you will not feel well.

6. You will have fevers, a slightly sore mouth, some skin rash and many severe stomach aches. You can also expect to vomit frequently, and lose some of your hair.

7. The disease, needles, strange surroundings and unknown people will make you feel afraid. You will be angry, or passive and withdrawn. This will be a fairly stressful time for your family.
CHANCE BOARD

NOTES:

1. lime green coloured translucent film overlay.

2. cut-out sections to display LIFE A numeric and graphical probability displays.

3. cut-out section for gripping probability wheel.

4. designers' initials (surnames).

5. dotted lines depict transparent plastic "parking permit" envelopes.

6. arrow suggests direction in which "wheel" be rotated.

7. background provided by white bristol board.

8. all lettering, arrows and lines are printed in black.

9. after all printing, colouring and trimming (including cut-outs) are complete laminate entire faceplate with transparent plastic film which is subsequently trimmed to one-eighth inch of faceplate edges.
Figure 2.2

LAYOUT OF CHANCE BOARD PROBABILITIES WHEEL

Scale 1 cm. = 1 inch.

- pink laminate or highlighter pen
  (contrasting pie - slices and probabilities coloured pale blue).
Figure 2.3

CHANCE (STANDARD GAMBLE) BOARD BACKPLATE SPECIFICATIONS

Scale: 1 cm = 1 inch

Notes:
1. backplate may be any colour or material but should be stiff and have sufficient strength to support the chance wheel using a grommet;
2. grommet is not centrally located;
3. no lettering or graphics are required;
4. lamination is optional and depends on material used.
Figure 2.4
FLOW DIAGRAM OF A CONVENTIONAL STANDARD GAMBLE QUESTION
(choice "A" is Perfect Health vs. Death, choice "B" is utility measurement state)

START

STEP 1
A: 100 vs. 0  B: inconsistent

STEP 2
A: 10 vs. 90  B: indifferent

STEP 3
A: 90 vs. 10  B

STEP 4
A: 20 vs. 80  B

STEP 5
A: 80 vs. 20  B

STEP 6
A: 30 vs. 70  B

STEP 7
A: 70 vs. 30  B

STEP 8
A: 40 vs. 60  B

STEP 9
A: 60 vs. 40  B

Preference for State B as compared to certain Death
UTILITY = 0.05, or = 0.00, or = neg.
SEQUENCE AND PLACEMENT OF HEALTH STATE DESCRIPTIONS FOR A SIMPLE CASCADE STANDARD GAMBLE EXAMPLE

QUESTION I:

QUESTION II:

QUESTION III:
TIME TRADE-OFF BOARD FACEPLATE SPECIFICATIONS

Scale: 1 cm = 1 inch.

Notes: 1. more preferred health state (often perfect health)
2. less preferred health state (transparent plastic envelope illustrated)
3. scale intervals (may be customized to application)
4. scale interval unit label (must be consistent with chosen scale
5. cut-out sections for gripping sliders
6. colours – background white, lettering black, health state cards yellow and laminated
7. bristol board faceplate laminated and trimmed to approximately one eighth
8. cut-out section to display "colour" intervals.
Figure 3.2

TIME TRADE-OFF BOARD BACKING PLATE SPECIFICATIONS

Scale: 1 cm = 1 inch

NOTES: 
1. black bristol board (laminated and trimmed to approximately one eighth inch of cardboard).
2. double-sided foam tape extending width of board (provides spacing and guidance for sliders plus adhering face and backing plates together).
3. cut-out sections for gripping sliders.
4. double-sided tape.
TIME TRADE-OFF BOARD SLIDER SPECIFICATIONS

- life A slider:

- life B slider:

Scale: $1 \text{ cm} = 1 \text{ inch}$. 
FIGURE 3.4

EXAMPLES OF TIME TRADE-OFF BOARD SCALES

ABSOLUTE SCALES

1. Pediatric - childhood only

2. Pediatric - expected lifetime

3. Young Adult - age 30 years

4. Geriatric - age 80 years

RELATIVE SCALES

5. Pediatric

6. Young Adult

7. Geriatric - age 70 years

8. Geriatric - age 80 years

SPECIALIZED RELATIVE SCALES

9. Two Stage (Geriatric)

10. Decreasing Precision (Geriatric)

11. Increasing Precision (Geriatric)

SCALES - Horizontal: 1 cm. = 1 inch.
Vertical: 1/2 cm. = 1 inch.
**Figure 3.5**

**EXAMPLE OF STEPS AND FORMATS FOR SIMPLE CONVENTIONAL TIME TRADE-OFF QUESTIONS**

<table>
<thead>
<tr>
<th>SERIAL STEP</th>
<th>OPTION</th>
<th>SCALE (YEARS OF LIFE REMAINING)</th>
<th>OUTCOME RESULTING</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>LIFE A</td>
<td><img src="image1" alt="Scale 1" /></td>
<td>Go to Step 2</td>
</tr>
<tr>
<td></td>
<td>LIFE B</td>
<td><img src="image2" alt="Scale 2" /></td>
<td>Stop: inconsistent answer (implies utility B &gt; utility A).</td>
</tr>
<tr>
<td>2.</td>
<td>LIFE A</td>
<td><img src="image3" alt="Scale 3" /></td>
<td>Stop: indifference &lt; 0, utility &lt; 0/6 = &lt; 0 (or go to Figure 3.6).</td>
</tr>
<tr>
<td></td>
<td>LIFE B</td>
<td><img src="image4" alt="Scale 4" /></td>
<td>Go to Step 3</td>
</tr>
<tr>
<td>3.</td>
<td>LIFE A</td>
<td><img src="image5" alt="Scale 5" /></td>
<td>Go to Step 4</td>
</tr>
<tr>
<td></td>
<td>LIFE B</td>
<td><img src="image6" alt="Scale 6" /></td>
<td>Stop: indifference = (6+5)/2 = 5.5 utility = 5.5/6 = 0.91.</td>
</tr>
<tr>
<td>4.</td>
<td>LIFE A</td>
<td><img src="image7" alt="Scale 7" /></td>
<td>Stop: indifference = (0+1)/2 = 0.5 utility = 0.5/6 = 0.08.</td>
</tr>
<tr>
<td></td>
<td>LIFE B</td>
<td><img src="image8" alt="Scale 8" /></td>
<td>Go to Step 5</td>
</tr>
<tr>
<td>5.</td>
<td>LIFE A</td>
<td><img src="image9" alt="Scale 9" /></td>
<td>Go to Step 6</td>
</tr>
<tr>
<td></td>
<td>LIFE B</td>
<td><img src="image10" alt="Scale 10" /></td>
<td>Stop: indifference = (5+4)/2 = 4.5 utility = 4.5/6 = 0.75.</td>
</tr>
<tr>
<td>6.</td>
<td>LIFE A</td>
<td><img src="image11" alt="Scale 11" /></td>
<td>Stop: indifference = (1+2)/2 = 1.5 utility = 1.5/6 = 0.25.</td>
</tr>
<tr>
<td></td>
<td>LIFE B</td>
<td><img src="image12" alt="Scale 12" /></td>
<td>Go to Step 7</td>
</tr>
<tr>
<td>7.</td>
<td>LIFE A</td>
<td><img src="image13" alt="Scale 13" /></td>
<td>Stop: indifference = (2+3)/2 = 2.5 utility = 2.5/6 = 0.42.</td>
</tr>
<tr>
<td></td>
<td>LIFE B</td>
<td><img src="image14" alt="Scale 14" /></td>
<td>Stop: indifference = (3+4)/2 = 3.5 utility = 3.5/6 = 0.58.</td>
</tr>
</tbody>
</table>
EXAMPLE OF A CARD DECK USED TO PRESENT STANDARD GAMBLE QUESTIONS

if respondent chooses this alternative "continue on to next card, CG".

if respondent is indifferent between alternatives "score is 0.9".

if respondent chooses this alternative "score is 0.925"

LEGEND:  - most preferred health state.
         - health state considered to be intermediate.
         - least preferred health state.

NOTE: percent of circle area covered by code is equal to percent chance of health type in gamble.
EXAMPLE OF A CHART AND MASK PROP USED TO PRESENT
TIME TRADE-OFF QUESTIONS

LEGEND: □ □ □ years of life remaining.
□ □ □ □ □ □ □ □ □ □ □ □ □ □ years of life lost to premature death.

NOTE: Dashed lines indicate lines not currently displayed (i.e., masked)
GLOSSARY OF TERMS
GLOSSARY OF TERMS

Accuracy
- The degree to which an estimate represents the true parameter.

Actuarial Life Expectancy
- The average number of additional years a person age X would live if current mortality trends continue to apply, based on the age-specific death rates for a given year.

Aesthetics
- The visual qualities of line, colour, form etc., that are pleasing or satisfying.

Anchor Bias
- Systematic error due to differences in the perceived limits (i.e., anchors) of a concept that results in a change in perceived context.

Attributes
- Dimensions or domains of health.

Backplate
- A piece of cardboard used to form part of a prop or visual aid.

Bias
- Results that depart systematically from the true values.

Bound (Upper and Lower)
- Abbreviation of boundary; represents the value of upper (maximum) and lower (minimum) limits of a parameter with defined confidence.

Cardinal
- Any number used in counting.
- A number from a measurement having interval or ratio scale properties.

Cascade
- A term used to describe a series of measures designed such that the results, when combined in a specific manner, will allow for the estimation of an implied response.

Chaining
- See "cascade".

Chart and Mask
- A type of prop or visual aid that consists of two cards, an overlay and an underlay.

Cognitive Burden
- The effort required to perceive, think, remember and make judgements.
Confidence Interval
- Consists of two boundary points between which we have a certain specified level of confidence that the population parameter lies (eg., 95% confidence interval for population mean).

Consent Form
- A printed document with blank spaces to be filled in by a respondent to provide written proof of consent to participate.

Conventional, Standard Gamble Question
- A standard gamble question in which the most desirable health state is perfect health and the least desirable state is death.

Converging Ping-Pong
- Technique that begins with extreme limit of interval (eg., duration of future life expectancy), and proceeds by alternating back-and-forth between high and low values while tending toward the central value of the interval.

D
- abbreviation for important, detectable difference.

Elicitation
- Drawing out or requesting.

Error term
- A random variable that describes how far away an individual response is from the population regression line.

Faceplate
- A piece of cardboard used to form the cover sheet of a visual aid (eg., Chance Board or Time Trade-Off Board).

Framing Effects
- A systematic change in a response due to variations in the perception of acts, contingencies, or outcomes. (Tversky and Kahneman, 1981)

Function
- Level of ability to perform a required or expected action.

Geriatric
- Old Age.

Homogeneity of Variance
- The variance of the dependent variable is the same for all values of the independent variable.

Implicit
- Suggested or to be understood though not plainly expressed.

Important
- Having much consequence or value.
Indifference Point
- Score or number that represents no preference between options.

Instrument
- A collection of tools for measurement of health-state utilities.

Interval
- The set containing all numbers between two given numbers and including one, both, or neither end point.
  (eg., Time Trade-Off Intervals)

Interval Scale
- A scale such that intervals of equal length are equal; that is the distance from 0.2 to 0.4 is the same as the distance from 0.4 to 0.6. Temperature scales are examples.

Interviewer Manual
- A handbook prepared by researchers as a guide and reference for interviewers.

Life Tables
- A summarizing technique used to describe the pattern of mortality and survival in populations.

Linear in Life Years
- The measure is a straight-line function of life expectancy.

Natural Scale
- A scale in which equal absolute amounts are represented by equal intervals.

Non-Verbal Prompts
- Directions that are not presented orally, for example: written or visual communication.

Ordinal Scale
- A scale such that values have a distinct order, or rank, but there is no natural distance between values. Social class is an example.

Pilot Testing
- Final trial prior to implementation of main study or survey (see prototype testing).

Power (i.e., Statistical Test Power)
- Probability of detecting a true difference of a specified size.

Precision
- The extent that a series of replicate observations of the same quantity conform to themselves (eg., may be expressed as standard deviation or standard error).

Prototype Testing
- Preliminary trials prior to pilot testing.

Range
- The absolute difference between the largest and smallest values in a defined interval.
Rank Preferences
- Ordinal scale preference measures in which health states are arrayed from high
to low (i.e., neglects distances between health states).

Rating Scale
- Same as visual analogue scale.

Ratio Scale
- An interval scale with a true zero point that permits ratios between values to
be meaningfully defined. Weight in kilograms is an example.

Relative Scale
- An interval with units that are meaningful only in relation to an individual’s
characteristic (e.g., percentage).

Respondent
- The individual providing answers to questions posed.

Response Booklets
- A set of forms used to identify, in code, the interview subject and record
answers to interview questions.

Scatter Diagram
- A graphic method of displaying the distribution of two variables in relation to
each other.

Schedule
- A written document that provides all details and directions required by the
interviewer to administer the instrument.

S.D.
- Abbreviation for standard deviation.

S.D. Diff
- Abbreviation for the standard deviation of the difference between means.

Significant
- Level of statistical confidence in a decision.

Slider
- A piece of cardboard or other material used to form part of a time trade-off
board.

Standard Gamble
- The classical method of measuring cardinal preferences, based on the fundamental
axioms of utility theory.

Statistically Independent
- Two events are independent if the occurrence or nonoccurrence of one of the
events does not change the probability of the occurrence of the other event.

Time Trade-Off
- An implicit method of measuring cardinal preferences that has been validated
against the standard gamble method.
Type I Error
- Rejecting a null hypothesis when it is actually true.

Type II Error
- Accepting a null hypothesis when it is incorrect.

Utility
- A cardinal (interval scale properties) preference measurement estimate that implies a notion of quality.

Utility Function
- A function relating the level of utility corresponding to a given set of levels for independent variables which affect utility, such as the level of consumption of goods, services, leisure and the health state. In health evaluation applications the utility function gives the relationship between the characteristics of the health state and the corresponding utility score for it.

Utility Score
- A cardinal preference observation for a health state as reported by an individual.

Validity
- An expression of the degree to which a measurement method measures what it is intended to measure.

Value
- An ordinal preference measurement estimate that implies a notion of quality.

Value Score
- An ordinal preference observation for a health state as reported by an individual.

Variability
- Amount of dispersion of the data.

Verbatim
- In exactly the same words.

Visual Analogue Scale
- A series of marks at graduated intervals along a line that is anchored at each end by a descriptive word or phrase describing the extreme limits of the scale (eg., best and worst).
DEFINITIONS IN GLOSSARY BASED ON INFORMATION FROM THE FOLLOWING SOURCES:


APPENDIX I
CONSENT FORM FOR PARENTS-OF-PATIENTS SURVEY

STUDY OF QUALITY OF LIFE FOR PATIENTS
TREATED FOR CHILDHOOD CANCER

I, __________________________, consent to participate in a study to determine parents' opinions about the quality of life experienced by children treated for cancer. The purpose of the study is to evaluate the quality of life in such patients both during and after treatment.

Mr./Mrs./Ms. __________________________ has explained to me that I will be asked to complete two 1 hour interviews at a time and place convenient to me.

I understand that there will be no direct benefit to me or my child from participating in this study and that my participation will in no way affect the treatment my child receives at the hospital.

I also understand that I may withdraw from the study at any time, even after signing this form. Any information that is collected about me during this study will be kept confidential and if the results are published, I will not be identified in any way.

Name (print) __________________________ Signature __________________________ Date __________________________

Witness (print) __________________________ Signature __________________________ Date __________________________

I have explained the nature of the study to the subject and believe he/she has understood it.

Name (print) __________________________ Signature __________________________ Date __________________________
CONSENT FORM FOR GENERAL PUBLIC SURVEY

STUDY OF QUALITY OF LIFE FOR PATIENTS TREATED FOR DISEASE DURING CHILDHOOD

I, ________________________, consent to participate in a study to determine how people feel about the quality of life experienced by children treated for disease. The purpose of the study is to evaluate the quality of life in such patients both during and after treatment.

Mr./Mrs./Ms. ______________________ has explained to me that I will be asked to complete two 1 hour interviews in my home at a time convenient to me.

I understand that there will be no direct benefit to me from participating in this study.

I also understand that I may withdraw from the study at any time, even after signing this form. Any information that is collected about me during this study will be kept confidential and if the results are published, I will not be identified in any way.

Name (print)  Signature  Date

Witness (print)  Signature  Date

I have explained the nature of the study to the subject and believe he/she has understood it.

Name (print)  Signature  Date
# INTERVIEWER MANUAL

FOR

CHILDHOOD HEALTH MEASUREMENT STUDY

## TABLE OF CONTENTS

### I. STUDY BACKGROUND

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Introduction and Purpose</td>
<td>85</td>
</tr>
<tr>
<td>2. The Researchers</td>
<td>87</td>
</tr>
<tr>
<td>3. Interview Consent</td>
<td>87</td>
</tr>
<tr>
<td>4. Publication and Confidentiality</td>
<td>88</td>
</tr>
<tr>
<td>5. Sample Cohort</td>
<td>88</td>
</tr>
<tr>
<td>6. Respondent Involvement</td>
<td>88</td>
</tr>
<tr>
<td>7. Pre-interview Contact</td>
<td>88</td>
</tr>
<tr>
<td>8. Interviewer's Role</td>
<td>89</td>
</tr>
</tbody>
</table>

### II. INTERVIEWING DETAILS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>9. Materials Required - Pre-Interview Check List</td>
<td>89</td>
</tr>
<tr>
<td>10. Interview Timing</td>
<td>90</td>
</tr>
<tr>
<td>11. Meeting Respondent(s)</td>
<td>91</td>
</tr>
<tr>
<td>12. Interview Setting and Family Members</td>
<td>91</td>
</tr>
<tr>
<td>13. Questionnaire Marking</td>
<td>92</td>
</tr>
<tr>
<td>14. Question Sequence</td>
<td>94</td>
</tr>
<tr>
<td>15. General Interviewing Considerations</td>
<td>95</td>
</tr>
<tr>
<td>16. Dealing with Inquiries and Pauses</td>
<td>95</td>
</tr>
<tr>
<td>17. Problem Questions</td>
<td>96</td>
</tr>
<tr>
<td>18. Commonly Asked Questions (with answers)</td>
<td>96</td>
</tr>
<tr>
<td>19. Checking the Questionnaire</td>
<td>100</td>
</tr>
<tr>
<td>20. End of Interview</td>
<td>100</td>
</tr>
<tr>
<td>21. Post-Interview Questions</td>
<td>100</td>
</tr>
<tr>
<td>22. Post-Interview Tasks</td>
<td>100</td>
</tr>
</tbody>
</table>
SECTION I: STUDY BACKGROUND

1. INTRODUCTION AND PURPOSE

This study has been designed by faculty members at McMaster University and physicians at both the McMaster University Medical Centre and the Hospital for Sick Children in Toronto. Funding has been provided by the Ontario Ministry of Health.

The purpose of this interview is to determine how parents evaluate factors that influence the state of children's health. We are interested in determining which characteristics (e.g., physical, emotional, mental, social) are most influential in making evaluations of long-term functional health. It must be emphasized that this interview is concerned only with the long-term functional health of children which must not be confused with illnesses or injuries that commonly affect children for short periods of time. Long-term problems present special needs which may require special treatments and training if children are to develop and adapt in such a way that they can successfully contribute to our community. The information that we are collecting over the course of this study will allow us to develop a system for measuring the functional health of children. This type of information can then be used to assess the needs of different groups of children and to evaluate the effectiveness of treatment programs.

The interviewing section of this study will consist of:

a) pretesting Stage I material;

b) Stage I survey (one interview per respondent);

c) pretesting Stage II materials; and

d) Stage II survey (one interview per respondent).

Pretesting of interview schedules was initially conducted by project staff and the
revised schedule pretested by two senior interviewers. All interviewers are expected to critique and familiarize themselves with the survey materials during the training sessions by alternating as interviewer and respondent.

Survey data will be collected from approximately 360 parents (290 general public and 70 parents-of-patients) during Stage I (Interview I) and 288 respondents (a subset of the original 360) in Stage II (Interview 2) at their place of residence. The parents-of-patients sample has been selected from children treated for childhood cancer at McMaster University Medical Centre or The Hospital for Sick Children. For the general public interviews we are asking interviewers not to mention the word "cancer" to the respondents. The use of the word "cancer" has been shown to predispose respondents attitudes or thinking when evaluating various health aspects during interviews and adversely influence their answers.

During the first interview (Stage I) each respondent will be asked to evaluate 7 different characteristics that influence children’s functional health. The first section involves a series of rating exercises in which the respondent is asked to rate levels of ability within each characteristic according to their personal preferences. Next the respondent will be asked to rank or order I3 health state descriptions based on combinations of characteristic levels. The third section requires the use of a chance board to evaluate four of the states. The last section of the first interview involves a series of personal questions which will allow us to categorize respondents by demographic and socio-economic factors.

The second interview (involving a subset of Interview I respondents) will require each respondent to evaluate a series of temporary or short-term health state descriptions.
These temporary health state descriptions describe some of the experiences of children during treatment for their disease.

2. THE RESEARCHERS

This study has been funded by the Ontario Ministry of Health and is being conducted by researchers at McMaster University. The principal investigators are:

(a) Dr. R. Barr;
(b) Dr. D. Feeny;
(c) Dr. P. Rosenbaum;
(d) Dr. G. Torrance;
(e) Dr. S. Weitzman; and
(f) Dr. A. Zipursky.

The project co-ordinator is Mr. Bill Furlong, the research assistant is Mr. John Horsman and secretarial assistance is provided by Miss Lori Johns. Staff assistance is available at McMaster University Medical Centre, Department of Clinical Epidemiology and Biostatistics (Telephone 525-9140 Ext. 2389).

3. INTERVIEW CONSENT

Consent is implicit when the respondent verbally agrees to be interviewed during the initial telephone contact. At the end of the first interview, the respondent is requested to sign a consent form stating that his/her participation is voluntary. Signed consent is necessary in order to satisfy the Ethics Committee of McMaster University. This consent does not imply that any names or personal data will be released and all information received during the interview must remain anonymous and confidential. The consent form is attached to the back of the Interview Response Booklet.
4. **Publication and Confidentiality**

No names or information from any one interview will be disclosed in any form. Data are presented only as applicable to groups of individuals (which is why socio-demographic questions are asked). If a respondent asks about the final report, inform them that the interviewing will not be completed until early 1987 and that the report preparations will probably take several more months.

5. **Sample Cohort**

Names of parents were identified from patient records (at McMaster Medical Centre and The Hospital for Sick Children) and student enrolment files. Names were drawn randomly from lists of students enrolled in Junior Kindergarten through Grade 5 in the City of Hamilton. Parents, matched by family and gender, of the randomly selected students will form the general public parent sample.

6. **Respondent Involvement**

Survey respondents will be asked to consent to participate in a series of two interviews. These interviews are to be conducted at the respondents residence and scheduled at the respondent’s convenience.

7. **Pre-Interview Contact**

Initial contact will be made by an announcement letter signed by Dr. R. D. Barr and Mr. Bill Furlong. Subsequently a telephone call will be made by an interviewer to:

a) check identity of Respondent (parent);

b) check eligibility (ability to speak and read English);

c) elicit a verbal parental consent to the series of parent interviews; and if possible
d) schedule the first interview in the parent's home.

8. INTERVIEWER'S ROLE

Interviewers play a major role in this study. All data will be collected by interviewers and therefore the quality of this data will be determined by the effectiveness of the interviewers. In general the effectiveness of an interviewer will depend on their ability to:

a) make a positive first impression on the Respondent to encourage their co-operation;

b) develop effective communication (at ease) with the Respondent to ensure their continued assistance;

c) avoid influencing the Respondent's opinion (may produce errors in the data set); and

d) collect and record (clearly and accurately) the responses obtained during the interview.

SECTION II: INTERVIEWING DETAILS

9. MATERIALS REQUIRED - PRE-INTERVIEW CHECK LIST

a) Interviewer Manual;

b) Introductory Letter (example copy);

c) Interview Schedule;

d) Response Booklet and Consent Form;

e) Feeling Thermometer;

f) 7 Sets of Health Characteristic Sticks (Communication X4, Physical/Mobility X5, Emotional X5, Learning X4, Self-care X4,
Pain/Discomfort X5, Fertility X3);
g) 14 Health State Cards (with attached sticks);
h) Standard Gamble Chance Board;
i) 4 Chance Cards (green coloured);
j) Income Card;
k) Eraser-tipped Pencils;
l) Tape recorder with two blank tapes; and
m) Spare tape recorder batteries.

**TASK CHECK-LIST**

Place cards and corresponding sticks together.
Place all cards and sticks in proper sections.
Label Response Booklet with respondent’s ID number.
Consent Form - print respondent’s name and your name in the blank on the form [I, respondent, consent to participate...

Mr./Mrs./Ms. interviewer has explained...].

Check tape recorder operation (install new batteries after every 3 interviews).

Label each tape with respondent’s ID number (same as Respondent Booklet).

Load tape, "side A" up, in the recorder.
Rewind to beginning of tape if necessary.

**10. INTERVIEW TIMING**

Punctual arrival will encourage co-operation with respondents. It is suggested that interviewers arrive in the respondents' neighbourhood five minutes prior to the scheduled interview time. Early arrival will provide time to:
a) clearly identify the specific residence;
b) park the car and walk to the building;
c) refresh your memory of respondent's names;
d) organize materials required for the interview; and
e) avoid the anxiety and dangers of rushing.

If you feel that you will be delayed don't hesitate to contact the scheduled respondent(s) to inform them of the situation.

11. MEETING THE RESPONDENT(S)

It is very important to place the respondent at ease as quickly as possible. First introduce yourself by name and study group (McMaster University). Next convey our appreciation for their agreeing to participate in this study and explain that obtaining their opinion is the only method available for determining what people feel are the most important factors affecting the long term health of children. At this time a few minutes should be spent to place the respondent at ease by introducing a casual conversation about a neutral topic such as the weather or a sporting event. When the interviewer senses that good communication has been established, they may introduce the requirements of the interview by proceeding with the Interview Schedule.

12. INTERVIEW SETTING AND FAMILY MEMBERS

The interview should take place in a private room with only the interviewer and respondent in attendance. Ideally the room should contain a fairly large table to support the thermometer scale, characteristic cards, board and response marking forms (a portion of the floor may have to be used).
It is important that the respondent be tactfully separated from family or friends so that only the designated person's opinions are being measured. Explain to the respondent and others that due to the selection process only one person's opinion can be accepted and must not be influenced by others. If the respondent should still question this procedure, explain that so far everyone has been interviewed individually, and it is important that the conditions be the same for all interviews. Some people may be hesitant to speak openly in front of others, so we designed the interviews to be conducted one at a time and privately.

13. QUESTIONNAIRE MARKING

Interview schedules and response booklets are different for general public and parent of patient respondents. The schedules and response booklets are identified as being the general public or parent-of-patient version on the cover sheet. The response booklets consists of:

1) an Interview Identification form (cover sheet);
2) a Characteristic Rating form (Table 1);
3) a Health State Rating form (Table 2);
4) a Chance Section form (Table 3);
5) a Demographic Section (pages 6 to 10); and
6) an Interview Assessment form (last page).

To ensure confidentiality the respondent's name must not appear on any interview recording materials. All marking must be done on the answer sheets in pencil. The Interview Response Booklet cover sheet will define each interview by Respondent, Interviewer and interview time. Each individual respondent will be identified by: 1) study identification number; and 2) gender (female or male). The codes associated with Respondent names appear only on the Respondent address list. The respondents sex
must be coded as a 1 for males and a 2 for females. The Interviewer codes are assigned by the field supervisor. The time and date should be recorded upon commencement of the interview.

The responses are to be recorded by the Interviewer. If during the interview an item or category is marked twice by mistake (e.g., more than one rating per item) the extraneous mark(s) should be erased completely before proceeding to the next question. In the event that the wrong marking technique is used (e.g., a check-mark instead of a circle), accept the form as is as long as one category is clearly assigned.

Characteristic rating scores are to be recorded by the Interviewer on Table 1. The score for each characteristic level, as indicated by the Respondent on the Feeling Thermometer, is to be recorded in the space provided adjacent to the respective level keywords. Scores for levels within each characteristic must not be recorded until the Respondent has completed any revisions they feel necessary.

Health state rating scores are to be recorded by the Interviewer on Table 2. The score for each state, as indicated on the Feeling thermometer, is to be recorded in the space provided adjacent to the respective keyword titles. Scores for each state must not be recorded until the Respondent has completed any revisions they feel necessary.

Scores derived from the three Chance questions may be read directly from the coding window of the Chance Board. In the event that the respondent chooses "Choice A" record the top value. If the respondent chooses "Choice B" record the bottom value but if the respondent indicates indifference between "A" and "B" record the centre value. In the event "Choice A" is selected and the top value is a "C". Change the "% chance" values by turning the wheel to the next setting as indicated in the Interview Schedule.
Similarly, for "Choice B", turn the wheel if the bottom value is a "C". If the value is a "?" read the question as indicated in the Interview Schedule and proceed as per instructions. "Chance" scores for the orange dot card are to be recorded in the upper section of Table 3. Scores for the yellow dot card must be recorded in the next section. Scores for the brown dot card are to be recorded in the next to last section while scores for the red dot card are to be recorded in the last section of Table 3 (page 5 of Response Booklet).

The schedule for the socioeconomic section is integrated with the response recording booklet. The Interviewer will read the questions contained in this section and record the Respondent's answers in the appropriate spaces.

Before taking leave of the Respondent, the Response Booklet must be checked for completeness and missing data recorded before the "end of interview" time is recorded.

After taking leave of the Respondent, the Interviewer will record their assessment of the Interview on the last page of the Response Booklet.

14. QUESTION SEQUENCE
The questionnaire has been designed in a specific order that must be maintained. In the Stage I interview the Respondent will first be asked to individually rate levels of 7 different characteristics. These characteristics must be presented to the Respondent one at a time as a set of sticks. The second section requires the Respondent to rank 14 health state descriptions (including death) in order of their preference or desirability. The 13 health state cards are to be presented sequentially, by section, A through D, to be followed by the death card. The third
section of the interview consists of four 'chance' type questions. The final section deals with personal or family characteristics and respondent interview evaluation.

15. GENERAL INTERVIEWING CONSIDERATIONS

a) Maintain a conversational tone throughout the interview by minimizing direct reading of the questionnaire. This is made easier if the interviewer is very familiar with the questions, the method of recording responses and the purpose of the survey.

b) Show interest in the study but do not react to the respondent's replies with any indication (verbal or facial expressions) of approval or disapproval. It is important to convey ACCEPTANCE to all respondent replies or the respondent may tailor responses to what the interviewer wants to hear.

c) Describe the study methods and ask the questions exactly as they appear on the schedule. BE CONSISTENT with all respondents. Paraphrasing is dangerous as this may change the meaning or emphasis of the question.

16. DEALING WITH INQUIRIES OR PAUSES

Try to encourage Respondents to rate or rank each item and answer every question. Use prompts when necessary but do not present the prompts in a leading manner or react to responses in an approving or disapproving manner. Some examples of suitable prompts are:

"If you had to choose one which would you choose?"

OR

"There are no right or wrong answers. What do you think?"

OR

"To the best of your knowledge, what would you say the answer would be?"
OR

"Take as much time as you wish, if you don't understand the question let me know and I will explain it to you."

17. PROBLEM QUESTIONS

All interviews are prone to disruptions resulting from threatening or embarrassing questions. In this survey, problem questions are most likely to occur in the final section dealing with socio-demographic issues. It is most important to maintain a good relationship with our volunteer Respondents, even at the expense of a missing response.

In the event that a Respondent appears uncomfortable with a question immediately relieve the tension by saying "Maybe you would like to think about this question, and we will come back to it later." If you make this statement before the Respondent actually refuses to answer the question you will:

a) retain the opportunity to re-ask the question at the end of the interview; and

b) ensure that the remaining portion of the interview is not placed in jeopardy by one question.

18. COMMONLY ASKED QUESTIONS

a) Why was I (or my family) chosen to be interviewed?

Two hundred and ninety Hamilton parents were chosen to take part in this survey. Children were chosen at random from lists of children who attend school in the City of Hamilton. The only criteria for being included in the study was that the children must have a parent of the same sex, living with the child, who is able to complete two interviews.
b) **Why are two interviews required?**

Two interviews are needed to maintain a practical interview duration and provide time between each set of questions. This time will be used to analyze the data and prepare the next set of questions.

c) **Why interview parents?**

Since this study is concerned with the relationships between medical care and children's health it is important to get the opinions of parents who are most often responsible for making health care decisions on behalf of their children.

d) **Why do you ask so many personal questions (e.g. employment)?**

These questions are asked for statistical purposes only. This information will allow us to divide the replies into groups. For example, we could compare all answers by people employed to all answers by people who are students. No information about any one person will be released and names are **not** recorded on the answer sheets.

e) **Will the results be used to measure my child's health status during this study?**

No, your answers to the questions in this study will only be used to develop a general measurement tool. In the second interview you will be asked to evaluate a different set of example cases but again these cases will not involve real people.

f) **What is the purpose of rating and ranking these factors?**

Results of the rating and ranking questions will provide the basic values used to form a mathematical model which will help to predict preference values for each health state.

g) **Why were these particular health factors chosen?**

These 7 characteristics were identified as important factors by health researchers and medical doctors. Each of these factors are common or important to personal health.

h) **What are the ages of the children to whom the characteristics refer?**

This study is concerned with evaluating characteristics of children between the ages
of 4 and 15 years.

i) What do you mean by "children's long-term function health"?

The children we are interested in are those people between the ages of 4 and 15 years. Children's long-term functional health includes all the things that affect the ability of children to function or do things in our community. In general this includes the children's physical condition, their mental abilities and their behaviour patterns but does not include short term illnesses such as coughs, colds or flu.

j. What do you mean by "a health index" or mathematical model?

A health index is a measurement tool used to determine a number or value which describes the long-term health of a child or a group of children.

k. Sometimes I find it really hard to make up my mind about an item. What should I do?

Take your time answering. Trust your instincts and answer as you feel. Don't worry about what other people may think or what the "score" should be. There are no right or wrong answers so simply give the best answer you can.

l) All aspects of health are important. How can I be expected to choose which is the most "preferable"?

Yes, all aspects of health are important but we feel that some levels of characteristics are likely to be considered more preferable than others. We are asking many different people to help us decide how much more or less preferable these levels are and which characteristics should be the focus of attention. Many people find it helpful to imagine that they are a child growing up today and make decisions based on how a child would decide on these questions. Parents have greater life experience and should trust their feelings in making judgments and remember that there are no right or wrong answers, just different opinions.
m) Why ask non-medical people to evaluate the factors that affect children's "well-being"?

It is important that we obtain opinions from members of the general public because they may have different values and experiences than health professionals. All members of the public have had some experience with handicapped individuals in their personal or family lives. These people can make valuable contributions in determining the important aspects of children’s well-being based on their day-to-day experiences. In addition to the practical value of the public's contribution to these questions this survey provides an opportunity for members of the public to influence the spending of public funds on medical treatment and research.

n) How often is "occasional, frequent and almost always"?

(The definitions of frequency of particular health state characteristics are contained in the Interview Schedule but should not be provided to the Respondent unless he/she specifically asks for them.)

- OCCASIONAL - 1 or 2 days a week or less
- FREQUENT - 3 or 5 days a week
- ALMOST ALWAYS - 6 or more days a week

o) What are "Normal Activities"?

Normal activities can simply be described as work, school or play.

p) Being blind is not the same as being deaf or mute. I would rate them differently if they were asked separately. How should I rate them?

Select the condition you feel is in the middle. In other words, rate the sticks by the condition you feel is neither the worst of the three nor the best.

(Ask which state is being rated and circle that condition on Table 1 of the Response Booklet.)
q) I find it difficult to imagine or think of myself as a 10 year old child while answering some questions, for example, about fertility!

Try to think of yourself as a child of 10 today. Do not think back to when you were 10 years old. You may find it easier to think of yourself as a 10 year old but with the knowledge of an adult and parent.

19. CHECKING THE QUESTIONNAIRE

Immediately after the interview, the interviewer should go through the entire set of questionnaire response forms to check that all responses are complete and legible. If a question has been overlooked or put-off to the end it must be completed before leaving the respondent. After ensuring that the questionnaire is complete the time should be recorded.

20. END OF INTERVIEW

At the end of the interview take time to express our appreciation for the time and effort the Respondent contributed to our research survey.

21. POST-INTERVIEW QUESTIONS

After leaving the Respondent complete the questions on the last page of the Response Booklet. These questions are an important part of evaluating the quality of each interview.

22. POST-INTERVIEW TASKS

Be sure all cards and sticks are back in their proper folders and that all sticks are with their proper cards and ready for the next interview.
APPENDIX III
CHILDHOOD HEALTH MEASUREMENT STUDY

GENERAL PUBLIC INTERVIEW SCHEDULE

START TAPE RECORDER - PUSH RED "RECORD" BUTTON

On behalf of the research team, I would like to thank you for agreeing to participate in this study. To begin, I will spend a few minutes explaining the purpose of this survey and why we feel it's important to have your assistance.

This study is being directed by a doctor in the Department of Pediatrics at the Chedoke-McMaster Hospitals in an effort to understand what people in Hamilton think about children's health. This information is important in helping to plan our current and future research and treatment activities.

All information is confidential and anonymous. Of course, your participation is voluntary, and if we should come to a question you would rather not answer, just let me know and we will skip it.

All of the questions deal with matters of opinion. There are no right or wrong answers, and everyone's opinions differ on these matters. You do not have to explain any of your answers, and your answers will not be questioned. All we want is your opinion.

The purpose of this interview is to obtain your opinions about long-term functional health of children. We are not concerned with illnesses like colds or flue that commonly affect children for short periods of time and from which they
readily recover. Some children have more abilities than others and some abilities may be considered more important than others. This study is concerned with a child's overall capacity to perform or do things over the course of their entire life. When rating these characteristics please imagine you are a child today, approximately ten years old, and spending the rest of your life in the conditions described. All decisions should represent what is best for you as a child rather than the parents or family.

In order to make this task a little easier we will use a Feeling Thermometer.

DISPLAY THERMOMETER

(place on table facing Respondent)

We call this a Feeling Thermometer because it helps us to measure people’s feelings about different things. In this case we are going to use the thermometer to measure your preferences for different health states.

This is how it works. The thermometer scale will indicate your preferences, from most desirable to least desirable. The more desirable or preferable you feel a health state to be, the closer it should be to the top of the thermometer.

RUN FINGER UP SCALE FROM 50 TO 100

The less preferable or desirable you think a health state is the closer it should be to the bottom.
Run finger down scale from 50 to 0

We are now ready to start. I will hand you cards with descriptions that include 7 aspects of health. In other words, each card consists of 7 statements and we want you to indicate your preference for each full description relative to every other description. The aspects of health are listed in the same order on each card but the level of abilities for one or more of the aspects of health will change on every card. There are a total of 13 of these health state cards.

Remove and show respondent the two cards from Section A

In order to make this task a little easier we have drawn coloured symbols beside sentences that describe less than perfect health. For example, on these two cards, one has no symbols and describes perfect health in all 7 categories. The other card has a symbol of different colours for each health category that indicates less than perfect health. In this case the symbols are squares which represent the lowest level of function ability in each category. There is a stick in the pocket on the back of the card and it has these same symbols and colours.

Remove sticks from both cards and hand to respondent

Each stick is divided into 7 sections which represent the 7 health categories. Please read each card and place the stick that describes perfect health on the left-hand side of the thermometer scale at 100. The card should be placed on the
right-hand side of the board across from the stick.

WAIT FOR RESPONDENT TO CHOOSE AND PLACE THE STICK AND CARD

(ADJUST IF NECESSARY)

The remaining card describes the lowest levels of health. As you can see, the colours, shapes and order of the symbols are the same on both the card and stick. The health category associated with each colour is printed on this stick as well. Place this stick at zero on the thermometer scale and its card at the lower right-hand edge of the board.

REMOVE 6 CARDS FROM SECTION B
WAIT FOR RESPONDENT TO CHOOSE, PLACE THE STICK AND CARD
(ADJUST IF NECESSARY)
HAND SET OF 6 CARDS TO RESPONDENT

Here are 6 more cards with sticks. Each describe health states that have only one health problem. The health problem is identified by the coloured symbol. Again the squares represent the lowest level of health. The other symbols represent less severe health problems. You may begin by reading the first card and placing its stick on the scale at the value that best represents your preference for that health state as compared to the two sticks already on the board. Place the card to the right of the scale opposite the stick. You may continue by reading the remaining cards, one at a time, and placing the sticks at locations on the scale that best represent your feelings as compared to the sticks already on the board.
WAIT FOR RESPONDENT TO COMPLETE THE 6 CARD SET

The cards are now becoming crowded at the edge of the board so I will shift the board slightly to allow space for the few remaining descriptions.

MOVE THE BOARD APPROXIMATELY 6 INCHES TO THE LEFT AWAY FROM THE CARDS TO ALLOW THE REMAINING CARDS TO BE PLACED SUCH THAT NO TWO CARDS OVERLAP

REMOVE 3 CARDS FROM SECTION C AND HAND THEM TO THE RESPONDENT

Each of these cards describe two health problems. Please read each one carefully and arrange the sticks beside the scale and the cards at the edge of the board. Please remember the distances between sticks should represent how much more preferable or less preferable you feel a health state to be when compared to the others. You may rearrange the order of the descriptions, move them further apart or closer together to suite your feelings as you consider these additional descriptions.

WAIT FOR RESPONDENT TO COMPLETE ALL 3 CARDS BEFORE SELECTING 2 CARDS FROM SECTION D (NOT 'DEATH' CARD) AND HANDING THEM TO THE RESPONDENT

These are the last two health state cards. These cards describe many different health problems. Please read each one carefully and place the sticks on the scale at values that best represent your preferences for these states when compared to the 11 states you have already evaluated. Again feel free to move the sticks and cards to best describe your feelings.
WAIT UNTIL RESPONDENT HAS PLACED ALL 13 STICKS ON THE SCALE

To complete this section we need one additional card. The card describes immediate death.

HAND RESPONDENT "IMMEDIATE DEATH" CARD FROM SECTION D

As we know everyone will die sometime although we don't know exactly what will cause it or when it will occur. If you feel it is always more preferable to live than to die, no matter how poor the functional health, place the stick at zero on the thermometer scale and move all the other health state descriptions up the scale to points that describe how much more preferable these states are when compared to death. On the other hand, if you feel that some functional health problems, as described on these cards, are worse than death, place death on the scale at the point that indicates how much more preferable death is to the states you feel are worse than death. Remember you may move any of the sticks to make room for the death state.

WAIT FOR RESPONDENT TO SCORE DEATH AND INDICATE THAT NO FURTHER CHANGES ARE PLANNED

Are you satisfied with your results? Now that you can see the results are there any changes you would like to make?
PAUSE UNTIL RESPONDENT INDICATES SATISFACTORY COMPLETION OF ANY REVISIONS, THEN RECORD SCORES FOR EACH CHARACTERISTIC ON TABLE #1 AS REPORTED BY RESPONDENT. (Page 2 of the Response Booklet)

We are now ready to record the values for each stick. On the back of each stick there is a label that corresponds to another label on the matching card.

DEMONSTRATE WITH BOTTOM STICK (AT 0) BEFORE RETURNING STICK AND CARD TO BOARD FOR RESPONDENT TO REPORT VALUE

In order to ensure that your scores are accurately recorded before the sticks are moved, would you first read the score for the least desirable state, then pick up that stick and read to me its identification label. Continue up the scale by first reading the score and then picking up each stick in turn and reading the label.

RECORD SCORES

Thank you. Would you please collect the cards and sticks and return the set to the white envelope. We are now ready to move onto the next section.

WHILE RESPONDENT COLLECTS STICKS AND CARDS:

- RECORD TIME AT BOTTOM OF TABLE 1.
- STOP RECORDER - TURN TAPE OVER IN RECORDER
- PUSH "RECORD" BUTTON TO START RECORDER

RETURN BINDER AND FEELING THERMOMETER TO CARRY-ALL BAG
CHANCE QUESTION

We have now finished with the Feeling Thermometer. For the next question I will show you 3 different descriptions of health and ask you to decide between two choices. If you think the two choices are equal, tell me. One choice will involve a risk and the other choice will be certain. The amount of risk will be changed until we find out how much risk you will take to avoid the certain choice. As before, there are no right or wrong answers, only what you think. In order to make the task easier to understand we will use an aid similar to a game board.

PLACE CHANCE BOARD ON TABLE, SET CHOICE A TO 90/10

TURN TO TABLE 2 OF RESPONSE BOOKLET (PAGE 3)

We call this a chance board because it indicates the chance or probability of a certain event occurring. As you can see the top part of the board is labelled Choice A.

POINT FINGER AT "CHOICE A"

The bottom part of the board is labelled Choice B.

POINT FINGER AT "CHOICE B"

You will be asked to pick Choice A or Choice B. Choice B, at the bottom of the board, will describe a state of health that is not the best or the worst of the three descriptions. Choice B is fairly simple because it describes only one form
of health and in the event that this option is chosen this form of health is certain to occur. If something is certain it is equal to a 100% chance or probability. However, Choice A is a little more complex because if it is chosen there are two possible results, perfect health and immediate death. The chances of each of these results occurring are indicated by the numbers appearing above each result and the size of the matching colours in the circle between the numbers. Another way of explaining the chance aspect of Choice A is that for every 100 patients who choose A, 90 will become perfectly healthy with treatment

**POINT TO 90**

but 10 will die during treatment

**POINT TO 10**

and none will know before choosing whether they will be one of the 90 or one of the 10. That is the chance they take. To make these ideas a little clearer let's run through a quick, imaginary example. Let us assume you have been involved in an accident and you have hurt your leg. When you see the doctor, he explains that you have two choices. Here are the descriptions for this example.

**PLACE EXAMPLE CARDS IN CHANCE BOARD POCKETS.**
**SET CHANCES TO 90/10**

Choice A is an operation and Choice B is to let your leg heal by itself. If you let the leg heal by itself, Choice B, it is certain that you will have a limp.
You will be able to walk, but you will not be able to run. On the other hand, you can choose the operation. The operation, Choice A in this example, is risky. It doesn't always work. If the operation does work, your leg is fixed and you can walk and run normally. If the operation does not work, you will have to use crutches.

The chance of walking and running normally after the operation is shown above the pink description and matches the proportion of pink colour in this circle.

---

**POINT TO PINK CHANCE VALUE AND PINK SECTION OF CIRCLE**

---

The chance of having to use crutches after the operation is shown above the blue description and matches the proportion of blue colour on this circle.

---

**POINT TO BLUE CHANCE VALUE AND BLUE SECTION OF CIRCLE**

---

The chances will change and I will ask you to choose Choice A or Choice B each time I change the chances.

---

**DEMONSTRATE BY TURNING WHEEL ON CHANCE BOARD**

---

Here, the chances of walking and running normally are 90% with a 10% chance of having to use crutches. If I were to spin this arrow, chances are it will probably land on pink because there is a very large amount of pink, 90 percent, showing in the circle. There is still a chance that it might land on blue, but it is a smaller chance, because there is only a small amount of blue showing on
the circle. Another way to think of it is that, on average, for every 100 people who choose A, the operation, 90 will walk and run normally afterwards but 10 will have to use crutches to get around after the operation.

CHANGE SPINNER AND BOARD TO 50/50

(IT IS SLIGHTLY QUICKER TO ROTATE THE WHEEL UPWARDS)

Now I’ve changed the chances. If I spin the arrow now, what colour is it most likely to stop on?

ALLOW RESPONDENT TO ANSWER

There is an equal chance that it will stop on pink or blue, because there is an equal amount of pink and blue showing on the circle. Do you understand how the Chance Board works?

IF ‘YES’, CONTINUE BELOW. IF ‘NO’, RETURN TO BEGINNING OF CHANCE SECTION, AND REPEAT THE EXERCISE.

Let’s work through the first question carefully together. In Choice A, the description will stay the same each time and is described by the pink card, Perfect Health, in the left-hand top pocket, and the blue card, Immediate Death, in the right-hand pocket. The health state of Choice B is one described by a green card and is one of the descriptions you evaluated using the Feeling Thermometer.
HAND RESPONDENT "ORANGE DOT" CARD

Please read over the description and when you are finished I will put it in pocket B at the bottom of the board.

PLACE "ORANGE DOT" CARD IN POCKET B WHEN RESPONDENT IS FINISHED
BE SURE WHEEL IS SET TO 100/0

We are now ready to begin. As you can see Choice A is now a 100% chance of perfect health and zero chance of death. Choice B is 100% chance of the health described on the card you just finished reading. I assume that you would pick Choice A, is that correct?

YES - MOVE WHEEL TO 10/90 AND CONTINUE OUTSIDE BOX BELOW
NO - CIRCLE RESPONSE (?), ASK "Why..." AND RECORD VERBATIM RESPONSE IN UPPER SECTION OF TABLE 3, PAGE 5.
- GO TO NEXT QUESTION.

Now I've changed Choice A to show that there is a 10% chance of perfect health and a 90% chance of death. Choice B is still a 100% chance of the health described on the card. Would you pick A or B now?
B - MOVE WHEEL TO 90/10, CONTINUE OUTSIDE BOX BELOW
A ?? - PROMPT "Do you mean you would prefer to have a 90% chance of dying immediately and a 10% chance of perfect health rather than living in the state of health described on the cards in Choice B?"
NO - REPEAT CHOICES SHOWN ON BOARD
YES - ASK "If Choice A was certain death with no chance of perfect health, do you think this would be better than living as Choice B, equal to living as Choice B or worse than living as Choice B?"
BETTER - MARK RESPONSE (NEGATIVE).
EQUAL - MARK RESPONSE (0.00).
WORSE - MARK RESPONSE (0.05).
STATE "Thank you that ends this question".
GO TO NEXT QUESTION.

The board now shows Choice A to be a 90% chance of perfect health with a 10% chance of dying and Choice B remains the same as before. Which choice would you now prefer?

A - MOVE WHEEL TO 20/80, CONTINUE OUTSIDE BOX BELOW.
B - MARK RESPONSE (0.95) AND STATE "Thank you that ends this question".
GO TO NEXT QUESTION.

Now I've changed Choice A to a 20% chance of perfect health with an 80% chance of dying immediately. Choice B is still a 100% chance of the health state described. Which choice would you prefer?

B - MOVE WHEEL TO 80/20, CONTINUE OUTSIDE BOX BELOW
A - MARK RESPONSE (0.15) AND STATE "Thank you that ends this question".
GO TO NEXT QUESTION.

Choice A is now an 80% chance of perfect health with a 20% chance of dying. Choice B is still the same as before. Would you prefer Choice A or B?
A - MOVE WHEEL TO 30/70, CONTINUE OUTSIDE BOX BELOW.
B - MARK RESPONSE (0.85) AND STATE "Thank you that ends this question".
    - GO TO NEXT QUESTION.

The choices have now been changed so that Choice A has a 30% chance of perfect health but a 70% chance of dying immediately. Choice B is still the same. Which choice would you now prefer?

B - MOVE WHEEL TO 70/30, CONTINUE OUTSIDE BOX BELOW.
A - MARK RESPONSE (0.25) AND STATE "Thank you that ends this question".
    - GO TO NEXT QUESTION.

Now I've changed Choice A to a 70% chance of perfect health and a 30% chance of immediate death. As before Choice B remains the same. Which choice would you now prefer?

A - MOVE WHEEL TO 40/60, CONTINUE OUTSIDE BOX BELOW.
B - MARK RESPONSE (0.75) AND STATE "Thank you that ends this question".
    - GO TO NEXT QUESTION.

Choice A has now been adjusted to indicate a 40% chance of perfect health and a 60% chance of immediate death. Would you prefer Choice A or B?

B - MOVE WHEEL TO 60/40, CONTINUE OUTSIDE BOX BELOW
A - MARK RESPONSE (0.35) AND STATE "Thank you that ends this question".
    - GO TO NEXT QUESTION.

If Choice A had a 60% chance of perfect health and a 40% chance of death as shown on the board would you pick choice A or B?
A - MOVE WHEEL TO 50/50, CONTINUE OUTSIDE BOX BELOW.
B - MARK RESPONSE (0.65) AND STATE "Thank you that ends this question".
- GO TO NEXT QUESTION.

Now I've changed Choice A to a 50% chance of perfect health and a 50% chance of immediate death. Choice B remains the same. Which choice would you now prefer?

A - MARK RESPONSE (0.45)
B - MARK RESPONSE (0.55)

Thank you, that ends this question.
APPENDIX IV
CHILDHOOD HEALTH MEASUREMENT STUDY

STAGE 1

GENERAL PUBLIC INTERVIEW RESPONSE BOOKLET

RESPONDENT I.D. /1-3

MALE/FEMALE /4

INTERVIEWER /5

START TIME: _______________ a.m./p.m. /6

INTERVIEW DATE: ________________, 1987 /7

CONFIDENTIAL (when completed)
### Table 1
**(Feeling Thermometer)**

<table>
<thead>
<tr>
<th>Health State</th>
<th>Thermometer Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Perfect Health</td>
<td>/8-10</td>
</tr>
<tr>
<td>TG</td>
<td>/11-13</td>
</tr>
<tr>
<td>SF</td>
<td>/14-16</td>
</tr>
<tr>
<td>UH</td>
<td>/17-19</td>
</tr>
<tr>
<td>RE</td>
<td>/20-22</td>
</tr>
<tr>
<td>VI</td>
<td>/23-25</td>
</tr>
<tr>
<td>OD</td>
<td>/26-28</td>
</tr>
<tr>
<td>WJ</td>
<td>/29-31</td>
</tr>
<tr>
<td>PC</td>
<td>/32-34</td>
</tr>
<tr>
<td>XK</td>
<td>/35-37</td>
</tr>
<tr>
<td>OB</td>
<td>/38-40</td>
</tr>
<tr>
<td>YL</td>
<td>/41-43</td>
</tr>
<tr>
<td>NA</td>
<td>/44-46</td>
</tr>
<tr>
<td>ZM</td>
<td>/47-49</td>
</tr>
<tr>
<td>Death</td>
<td>/50-52</td>
</tr>
</tbody>
</table>

Table 1 finished at __:__ am/pm
<table>
<thead>
<tr>
<th>TABLE 2</th>
<th>(CHANCE BOARD)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>ORANGE DOT CARD</strong></td>
<td></td>
</tr>
<tr>
<td>RECORD SCORE INDICATED ON BOARD = <em><strong>.</strong></em> ___</td>
<td>/1-3</td>
</tr>
<tr>
<td>OR</td>
<td>/4</td>
</tr>
<tr>
<td>? - PROMPT &quot;Why did you choose a 100 percent chance of health B rather than a 100 percent chance of perfect health?&quot;</td>
<td>/5</td>
</tr>
<tr>
<td>RECORD VERBATIM RESPONSE:</td>
<td>/6</td>
</tr>
<tr>
<td>&quot;Thank you, that ends this question&quot;. GO TO NEXT QUESTION.</td>
<td>/7</td>
</tr>
</tbody>
</table>

| **YELLOW DOT CARD**          |                |
| RECORD SCORE INDICATED ON BOARD = ___.___ ___               | /8-11          |
| OR                           | /12-15         |
| ? - PROMPT "Why did you choose a 100 percent chance of health B rather than a 100 percent chance of perfect health?" | /12-19         |
| RECORD VERBATIM RESPONSE:    | /16-19         |
| "Thank you, that ends this question". GO TO NEXT QUESTION. |                |
TABLE 2 continued

RED DOT CARD

RECORD SCORE INDICATED ON BOARD = ___.___ ___

OR

? - PROMPT "Why did you choose a 100 percent chance of health B rather than a 100 percent chance of perfect health?"

RECORD VERBATIM RESPONSE:

________________________________________________________________________________________

"Thank you, we are now finished with the Chance Board."

Table 2 finished at ___:___ am/pm

GO TO "FAMILY QUESTIONS" - NEXT PAGE OF THIS BOOKLET.
FAMILY QUESTIONS

This is the last section of the interview. It concerns family related information and these questions will allow us to divide respondents into groups. As with everything in this interview, your name will not be recorded on this form. All information is confidential and anonymous.

1. How many children do you have?_____

2. Do any of your own children currently have any major medical conditions or health problems of a permanent or long term nature?

DEFINITION OF MAJOR HEALTH PROBLEM:

"A health problem seen by the parent as major; a long-term condition; a problem which may have long-term effects on the child's life."

1. Yes
2. No
7. Refused
8. Don't know

3. Do any of your children currently have any major problems with school?

DEFINITION OF "MAJOR PROBLEMS WITH SCHOOL":

"A problem with school that the parent defines as major. The problem may be a learning disability, deafness, blindness, a speech problem or a behavioural handicap such as an emotional or disciplinary problem."

1. Yes
2. No
7. Refused
8. Don't know
4. Do you know by name any handicapped children, such as children in your
neighbourhood, family members, or friends?

( ) 1. Yes
( ) 2. No
( ) 7. Refused
( ) 8. Don’t know

5. How would you describe your present state of health? Would you say its
excellent, good, fair, poor?

( ) 1. Excellent
( ) 2. Good
( ) 3. Fair
( ) 4. Poor
( ) 7. Refused
( ) 8. Don’t know

6. What is your marital status?

( ) 1. Single
( ) 2. Married or Common Law
( ) 3. Divorced or Separated
( ) 4. Widowed
( ) 5. Other (specify: ________________________)
( ) 7. Refused

7. What is your date of birth?

YEAR    MONTH    DAY
8. What is the highest level of formal education you achieved?

01 No formal schooling (self taught) ____________________________ ( )
02 Some elementary schooling ________________________________ ( )
03 Completed elementary schooling ____________________________ ( )
04 Some secondary schooling _________________________________ ( )
05 Secondary School graduation certificate ____________________ ( )
06 Apprenticeship or Journeyman ______________________________ ( )
07 Non-university certificate or diploma (e.g. College diploma, etc.) ____________________________ ( )
08 Some university experience ________________________________ ( )
09 Bachelor's degree(s) ____________________________________ ( )
10 Degree in medicine, dentistry or veterinary medicine (M.D., D.D.S. or D.M.D. or D.V.M.) ____________________________ ( )
11 Master's degree(s) ______________________________________ ( )
12 Earned doctorate (e.g. Ph.D.) ______________________________ ( )
13 Other (specify): _________________________________________ ( )
77 Refused ________________________________________________ ( )
88 Don't Know ____________________________________________ ( ) /37-38

9. What is your religion?

( ) 01 Roman Catholic
( ) 02 United Church
( ) 03 Anglican
( ) 04 Presbyterian
( ) 05 Other Protestant
( ) 06 Eastern Orthodox
( ) 07 Jewish
( ) 08 Eastern Non-Christian
( ) 09 No Religious Preference
( ) 10 Other (specify) ________________________________
( ) 77 Refused
( ) 99 Don't Know /39-40
10. On average how often do you participate in a religious service?

( ) 1  Once or more per week
( ) 2  One to three times per month
( ) 3  Less than once a month
( ) 4  Do not participate
( ) 5  Refused
( ) 6  Don't Know

11. Could you tell me which letter on this card...

HAND RESPONDENT INCOME CARD (Pocket on Back Cover of SCHEDULE)

best represents your approximate total family income for 1986? Please include income from all sources such as wages, salaries, pensions, family allowances, rents from properties, and so forth. (If Respondent seems uncertain, probe for best estimate possible).

( ) 1. A (under $12,000)
( ) 2. B ($12,000 - $17,999)
( ) 3. C ($18,000 - $23,999)
( ) 4. D ($24,000 - $29,999)
( ) 5. E ($30,000 - $34,999)
( ) 6. F ($35,000 - $39,999)
( ) 7. G ($40,000 and over)
( ) 8. Refused
( ) 9. Don't know --- PROBE: If you had to guess, what would you say?
TEAR OFF AND HAND RESPONDENT CONSENT FORM
ATTACHED TO BACK OF THIS RESPONSE BOOKLET

For our records, would you sign this form stating that your participation was voluntary?

AFTER RESPONDENT HAS SIGNED THE CONSENT FORM, RETRIEVE AND SIGN IT.

The final questions ask for your opinions of this interview. We constantly try to improve our method of collecting information. Any assistance you may provide would be appreciated.

12. How would you rate the Feeling Thermometer and questions?........
   ( ) 1. Very easy to understand
   ( ) 2. Easy to understand
   ( ) 3. Neither easy nor difficult
   ( ) 4. Difficult
   ( ) 5. Very difficult
   ( ) 7. Refused
   ( ) 8. Don't Know

13. How would you rate the Chance Board and its questions?.......  /43
   ( ) 1. Very easy to understand
   ( ) 2. Easy to understand
   ( ) 3. Neither easy nor difficult
   ( ) 4. Difficult
   ( ) 5. Very difficult
   ( ) 7. Refused
   ( ) 8. Don't Know

   Thank you very much. Is there anything else you would like to say or add about the interview?

/45-4
Well this ends the interview. Before scheduling the next interview we must analyze the results of these interviews. We expect to contact you in about six weeks but it may take a little longer and require a special effort to avoid conflict with your holiday plans. If you are planning to be away this summer it would be very helpful if I knew which dates to avoid when attempting to schedule the next interview. Are you planning to be away from home?

( ) NO
( ) YES

HOLIDAY DATES:

FROM: ___________ TO: ___________
FROM: ___________ TO: ___________
FROM: ___________ TO: ___________
FROM: ___________ TO: ___________

Thank you for your co-operation and assistance. I’ve enjoyed talking to you and hope you’ve found this a pleasant experience.

TIME INTERVIEW ENDED:

__________________ a.m./p.m.

TOTAL TIME FOR INTERVIEW

__________________ MINUTES

/48-5C
TO BE COMPLETED AFTER THE INTERVIEW

17. Did anyone other than the Respondent contribute information?
   ( ) 1. YES PLEASE EXPLAIN ________________________________
          ________________________________ ________________________________
   ( ) 2. NO ________________________________________________ /5
          _____________________________________________________________

18. Rate degree of co-operation from Respondent in following categories:
   ( ) 1. Complete co-operation /5%
   ( ) 2. General co-operation -- not fully open on all questions
   ( ) 3. Substantial lack of co-operation

19. How much thought did the Respondent put into his/her answers?
   ( ) 1. A great deal /5:
   ( ) 2. Some
   ( ) 3. Very little
   ( ) 4. None at all

20. How well did the Respondent understand the questions?
   ( ) 1. Totally /5:
   ( ) 2. For the most part
   ( ) 3. Somewhat
   ( ) 4. Only a little
   ( ) 5. Not at all

21. How much trouble did the Respondent have in answering these questions?
   ( ) 1. A great deal /5
   ( ) 2. Some
   ( ) 3. Very little
   ( ) 4. None at all

22. Record impression of the interview (i.e., the quality of response):
   ( ) 1. Very good COMMENTS: ________________________________
   ( ) 2. Good ________________________________________________ /5
   ( ) 3. Average ______________________________________________
   ( ) 4. Poor ________________________________________________
   ( ) 5. Very poor ______________________________________________
Table 2
CHANCE BOARD

**RECORD 3 HEALTH STATE CODES

<table>
<thead>
<tr>
<th>THERMOMETER RANKING</th>
<th>CARD PILE SEQUENCE</th>
<th>CARD PLACEMENT FORM</th>
<th>(from TABLE 1)</th>
<th>(l/line)</th>
<th>(face-up)</th>
<th>Question - Pocket</th>
</tr>
</thead>
<tbody>
<tr>
<td>TOP</td>
<td>I</td>
<td>GREEN</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CENTRE</td>
<td>I</td>
<td>BLUE</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BOTTOM</td>
<td>II</td>
<td>GREEN</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>III</td>
<td>BLUE</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

DEATH (turn over 'DEATH' card -------> III ---> BLUE in BLUE pocket)

QUESTION I: SCORE = __________
OR
? - PROMPT "Why did you choose a 100 percent chance of health B rather than a 100 percent chance of perfect health? RECORD VERBATIM RESPONSE:

<table>
<thead>
<tr>
<th>Perfect ** Health **</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
</tbody>
</table>

QUESTION II: SCORE = __________
OR
? - PROMPT "Why did you choose a 100 percent chance of health B rather than a 100 percent chance of perfect health? RECORD VERBATIM RESPONSE:

<table>
<thead>
<tr>
<th>Perfect ** Health **</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
</tbody>
</table>

QUESTION III: SCORE = __________
OR
? - PROMPT "Why did you choose a 100 percent chance of health B rather than a 100 percent chance of perfect health? RECORD VERBATIM RESPONSE:

<table>
<thead>
<tr>
<th>Perfect Health Death **</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
</tbody>
</table>

TABLE 2 FINISHED AT _____:____ am/pm

CASCADE VALUES - TABLE 2, VERSION B
SA = __.__
DQ = __.__
MU = __.__

/41- /44-
/47-
TIME TRADE-OFF SCHEDULE-EXAMPLE

In each question I will present to you a series of two choices and ask you to choose the one that you would prefer. If you think the two choices are equal tell me and I will mark them both. In order to make the task easier to understand we will use an aid similar to a game board.

PLACE TIME TRADE-OFF BOARD ON TABLE
SET STAR A TO 5 YEARS AND STAR B TO 6 YEARS
PLACE YELLOW CARD #62 IN SPACE B
TURN TO TABLE 1 OF RESPONSE BOOKLET (for example see Appendix VII)

As you can see the top part of the board is labelled life A and the bottom half of the board is labelled life B. These are your two choices. The cards

POINT TO BOTH CARDS

describe the functional health of each life. The time scales beside each card show the time of death and the number of years lost to life due to an early death.

Please remember that we are interested in the functional health during the remaining years of life and therefore, we have started the time scale at zero.

POINT TO 0 ON SCALE A

The scale extends to 6 years which represents the average expected years of life for an eighty year old male.

POINT TO 6 ON SCALE A
The red colour on the scale shows the years of excellent functional health.

**RUN FINGER ALONG RED PART OF SCALE A**

Age of death is marked by a star.

**POINT TO STAR ON SCALE A**

The years lost to life due to an early death are shown by the black colour.

**POINT TO BLACK PORTION OF SCALE A**

Do you understand these ideas?

**YES - SET STARS A AND B TO 6 YEARS AND CONTINUE BELOW**

**NO - REPEAT PREVIOUS PAGE.**

Please read both cards A and B.

**WAIT FOR RESPONDENTS TO INDICATE COMPLETION**

Let's start the first question by working through it together. The top part of the board represents life A. The card describes a healthy life.

**POINT**
The time scale, marked in red

POINT

shows that good health, will last throughout life until death. Death is marked by a star

POINT

and would occur after 6 years. The bottom card and time scale describe life B. In this case the card describes some health problems. The time scale on the right, marked in blue

POINT

shows that these problems will continue for 6 years until death.

POINT

I assume that you would prefer life A. Is that correct?

A - MARK RESPONSE

- MOVE STAR A TO 0 YEARS AND CONTINUE OUTSIDE BOX BELOW

B - MARK RESPONSE, ASK "WHY" AND RECORD VERBATIM RESPONSE

Now I'll change the life A time scale to show immediate death. Which life would you choose now?
Now I have made life A five years long with one year lost due to premature death.

Which life would you now prefer?

Life A is now 1 year long with 5 years lost. Now which life would you prefer?

If life A was 4 years long with 2 years of lost life which would you choose?

Life A now ends with death after 2 years and 4 years of life have been lost. Which life would you prefer?
A - MARK RESPONSE, GO TO QUESTION 2 OF THIS SCHEDULE.

B - MARK RESPONSE, MOVE STAR A TO 3 YEARS AND CONTINUE BELOW.

Life A is now 3 years long with 3 years of life lost. Which would you choose now?

A AND/OR B - MARK RESPONSE(S), GO TO QUESTION 2 OF THIS SCHEDULE
TIME TRADE-OFF RESPONSE MARKING FORM

TABLE 1

(TIME TRADE-OFF BOARD, CASE CARD 62)

<table>
<thead>
<tr>
<th>Life A</th>
<th>Life B</th>
</tr>
</thead>
<tbody>
<tr>
<td>( ) 6</td>
<td>6 ( )</td>
</tr>
<tr>
<td>STOP</td>
<td>( ) 0&lt;</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>( ) 5</td>
<td>6 ( ) STOP</td>
</tr>
<tr>
<td>STOP</td>
<td>( ) 1&lt;</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>( ) 4</td>
<td>6 ( ) STOP</td>
</tr>
<tr>
<td>STOP</td>
<td>( ) 2&lt;</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>( ) 3&lt;</td>
<td>6 ( ) STOP</td>
</tr>
</tbody>
</table>

PROMPT:
"Why did you choose 6 years of health problems?"

RECORD VERBATIM RESPONSE:

______________________________
______________________________
______________________________

STOP
APPENDIX VIII
LISTS OF MATERIALS FOR CONSTRUCTING VISUAL AIDS

A. FEELING THERMOMETER

1. visual scale - graduated intervals from 0 to 100
   - 9 cm. by 60 cm.
   - high contrast photographic line print on resin coated paper.

2. felt - 65 cm. by 25 cm. of standard, fabric store felt.

3. board - 65 cm. by 40 cm. of white foamboard, arbourite or other suitable material.

4. pointers or arrows
   - size varies with application.
   - fabricated from balsa wood, plastic laminated paper or other suitable materials.

B. CHANCE BOARD

1. white bristol board (cardboard)
   - 3 pieces: 33 cm. by 33 cm.
     : forms faceplate, probabilities wheel, backplate.

2. plastic lamination film
   - protects faceplate, probabilities wheel and backplate.

3. graphics (i.e., letters and numerals)
   - self-adhesive graphics such as Geotype:
     304-24 Helvetica Medium (approx. 6 mm.); and
     303-48CN Helvetica Light (approx. 13 mm.).

4. three transparent plastic holders for health-state cards
   - (eg., parking permit holders)

5. graphic art, transparent, colour/tint overlay film
   - pink (eg., Letraset PANTONE 243-A);
   - blue (eg., Letraset PANTONE 290-A); and
   - green (eg., Letraset PANTONE 351-A).

C. TIME TRADE-OFF BOARD

1. white bristol board (cardboard)
   - 1 piece: 33 cm. by 33 cm.
     : forms faceplate.
   - 2 pieces: 33 cm. by 6 cm.
     : form sliders.

2. black bristol board (cardboard)
   - 1 piece: 33 cm. by 33 cm.
     : forms backplate.
3. plastic lamination film
   - protects faceplate, sliders and backplate.

4. graphics (i.e., letters and numerals)
   - self-adhesive graphics such as Geotype:
     304-24 Helvetica Medium (approx. 6 mm.); and
     303-48 CN Helvetica Light (approx. 13 mm.).

5. two, transparent plastic holders for health-state cards (eg., parking permit holders, bank book envelopes etc.).

6. graphic art, transparent, colour/tint overlay film for sliders:
   - pink (eg., Letraset PANTONE 243-A);
   - blue (eg., Letraset PANTONE 290-A); and
   - black (eg., Letraset PANTONE Extra Black A, or exposed photographic paper etc.).