THE EFFECT OF AMBIGUITY IN HEALTH STATE DESCRIPTIONS
ON THE INTERNAL CONSISTENCY OF
STANDARD GAMBLE UTILITIES

by

Dena M. Bravata
Ingram Olkin
Richard W. Cottle
B. Curtis Eaves
Alan M. Garber
Mary K. Goldstein

May 2000

Department of Statistics
STANFORD UNIVERSITY
Stanford, California
THE EFFECT OF AMBIGUITY IN HEALTH STATE DESCRIPTIONS
ON THE INTERNAL CONSISTENCY OF
STANDARD GAMBLE UTILITIES

by

DENA M. BRAVATA
INGRAM OLKIN
RICHARD W. COTTLE
B. CURTIS EAVES
ALAN M. GARBER
MARY K. GOLDSTEIN

MAY 2000

Financial support for this study was provided in part by a grant from the
National Institutes for Health grant AG15110

Department of Statistics
Dequioia Hall
STANFORD UNIVERSITY
Stanford, California 94305-4065

http://www-stat.stanford.edu
The Effect of Ambiguity in Health State Descriptions on the Internal Consistency of Standard Gamble Utilities


The manuscript is not under review elsewhere, the primary data have not been published previously or accepted for publication, and the appropriate ethical guidelines were followed in the conduct of the research. Portions of this work have been presented at the Twenty-first Annual Meeting, Society for Medical Decision Making; Reno, Nevada; October, 1999 and the Annual Meeting of the American Geriatrics Society; Nashville, Tennessee; May, 2000.

Financial support for this study was provided in part by a grant from the National Institutes for Health grant AG15110. Dr. Bravata is a VA Ambulatory Care Fellow, VA Health Care System, Palo Alto, CA. Dr. Goldstein is supported by a Career Development Award from the VA Health Services Research and Development Service. The funding agreement ensured the authors’ independence in designing the study, interpreting the data, writing and publishing the report. We have no other financial interests or other dual commitments that might represent potential conflicts of interest.

Please address all communication concerning the manuscript to: Dena M. Bravata, M.D., M.S., Center for Primary Care and Outcomes Research, Stanford University School of Medicine, 179 Encina Commons, Stanford, CA 94305-6019, Telephone: 650-723-1118; Fax: 650-723-1919; E-mail: bravata@healthpolicy.stanford.edu.
The Effect of Ambiguity in Health State Descriptions on the Internal Consistency of Standard Gamble Utilities

Dena M. Bravata, M.D., M.S., 1,2 Ingram Olkin, Ph.D., 3,4 Richard W. Cottle, Ph.D., 5
B. Curtis Eaves, Ph.D., 5 Alan M. Garber, M.D., Ph.D., 1,2 Mary K. Goldstein, M.D., M.S. 1,2

AUTHOR AFFILIATIONS

1VA Palo Alto Health Care System, Palo Alto, CA. 2Center for Primary Care & Outcomes Research, Department of Medicine, Stanford University School of Medicine, Stanford, CA.
3Department of Statistics, Stanford University, Stanford, CA. 4School of Education, Stanford University, Stanford, CA. 5Department of Management Science and Engineering, Stanford University, Stanford, CA.

RESEARCH PERFORMED

June 1999 to June 2000
GRANT SUPPORT

Dr. Bravata is a VA Ambulatory Care Fellow, VA Health Care System, Palo Alto, CA. Dr. Olkin is supported by National Science Foundation grant DMS 96 26 265. Dr. Goldstein is supported by a Career Development Award from the VA Health Services Research and Development Service. Major support for this research is from National Institutes on Aging grant AG15110.

CORRESPONDING AUTHOR

Dena M. Bravata, M.D., M.S.
Center for Primary Care and Outcomes Research
Stanford University School of Medicine
179 Encina Commons
Stanford, CA 94305-6019
Voice: 650-723-1118
Fax: 650-723-1919
E-mail: bravata@healthpolicy.stanford.edu
ABSTRACT

The unavoidable ambiguity of health state descriptions can affect the frame of reference adopted by subjects as they respond to utility questions. Support theory describes the two extreme frames of reference: the minimal and the inclusive. We sought to develop and evaluate approaches for validating the consistency of utilities accounting for differences in interpretations of health states according to the frame of reference adopted by the subject. We used the standard gamble technique to elicit utilities for health states characterized by dependence in the following activities of daily living (ADLs): dressing, walking, bathing, continence, toileting, eating, and transferring. We evaluated each subject’s utilities for inconsistencies given minimal and inclusive frames of reference and investigated differences in utilities between consistent and inconsistent subjects.
Standard Gamble Utilities in Health Care

Utilities are the standard measure of value in the analysis of health decisions (Chapman & Sonnenberg, 2000; Gold, Siegel, Russell, & Weinstein, 1996). Among the various methods for eliciting utilities, the standard gamble technique is favored for being grounded in a normative theory of preference with well-described mathematical and theoretical foundations (Edwards, 1992; Fishburn, 1982; von Neumann & Morgenstern, 1944). In the standard gamble, a respondent chooses between two alternatives, one with a certain outcome and one that involves a gamble. Von Neumann and Morgenstern first proposed the “reference gamble” or “reference lottery” in the context of game theory (von Neumann & Morgenstern, 1944). Schlaifer later described the “standard gamble” in the context of decision making for business administration as the combination of multiple reference gambles when decision makers make judgments about alternative outcomes separately rather than simultaneously (Schlaifer, 1969). Torrance and colleagues were among the first to use the standard gamble to elicit utilities for the analysis of medical decision making (Torrance, Thomas, & Sackett, 1972).

The current technique for using the standard gamble to elicit utilities for health states begins with asking the respondent to consider a hypothetical choice between the certainty of continued life in the health state of interest (a health state of less than optimal health) and a gamble (Gold et al., 1996; Torrance, 1986; Torrance & Feeny, 1989; Weinstein, 1980). The gamble has two possible outcomes (Figure 1). The best outcome is usually a state of perfect health (assigned a utility value of 1), which is described as occurring with probability \( p \). The worst outcome is usually death (assigned a utility value of 0), which is described as occurring with probability \( 1 - p \). The probabilities in the gamble are adjusted until the respondent is indifferent between the choice of the certainty of continued life in the health state of interest and the gamble. The expected value of the gamble, by substitution, is the utility for the health state of interest relative to full health and death. This process is repeated for all health states to be rated.
For the purposes of this paper, we assume that the standard gamble is used to elicit utilities for health states that are better than death with the anchors of perfect health and death. A better-than-death health state is a state in which longer survival is preferred to shorter survival (Miyamoto, 2000). Not all health states are better than death. Such health states give rise to important and interesting assessment issues, but a discussion of these issues would digress from the central questions of this paper (Llewellyn-Thomas et al., 1984; Patrick, Starks, Cain, Uhlmann, & Pearlman, 1994).

**Cognitive Biases of the Standard Gamble**

Despite the widespread acceptance of the standard gamble and its theoretical grounding, extensive empirical research has demonstrated a variety of ways in which preferences are inconsistent with the assumptions of expected utility theory (Kahneman & Tversky, 1979; Slovic, Lichtenstein, & Fischhoff, 1988). Most inconsistencies in preferences for health states that result from limitations in human judgment arise when the same objective alternatives are viewed in relation to different points of reference (Dolan & Kind, 1996; Froberg & Kane, 1989). Kahneman and Tversky called these inconsistencies “framing effects” (Kahneman & Tversky, 1981). Their research and that of others have demonstrated that standard gamble utilities differ when the gamble is presented in a “positive” frame that emphasizes the chance of survival (e.g., “80% chance of survival”) versus a “negative” frame that emphasizes the risk of dying (e.g., “20% risk of death”) (Eraker & Sox, 1981; Kahneman & Tversky, 1981; Llewellyn-Thomas et al., 1982; McNeil, Pauker, Sox, & Tversky, 1982; Percy & Llewellyn-Thomas, 1995). This reversal of preferences occurs despite the fact that the two outcomes are identical. Preferences for health states should not change with changes in frame, just as the height of two neighboring mountains should not reverse with changes in vantage point.

According to normative decision theory, the likelihood of an event (such as a diagnosis or treatment outcome) should be expressed as a probability, and probability judgments should obey the laws of probability theory. The rules of probability apply even when the likelihood is an
intuitive estimate (Chapman & Elstein, 2000; Hagen, 1991; Keeney & Raiffa, 1993; Nau, 1981; Weinstein, 1980). For the purposes of our consistency evaluation, we assume that multiattribute utilities are joint probabilities (Keeney & Raiffa, 1993). For example, a subject who reports willingness to risk a 10 percent chance of death to be cured of a given health state would be considered to have a utility of 0.9 for that health state and a probability of 0.9. We will refer to these joint probabilities simply as expected utilities, denoted EU.

Tversky and Koehler developed support theory to explain the observation that alternative descriptions of the same outcome can give rise to different probability judgments (Tversky & Koehler, 1994). The central idea of support theory is that probabilities are assigned to descriptions of events, not to the events themselves. For example, the explicit description “death due to traffic accident, drowning, electrocution, or any other accident” and the implicit description “death due to accident” represent different descriptions of the same event for which subjects may have different preferences (Redelmeier, Koehler, Liberman, & Tversky, 1995). Tversky suggested that people generally evaluate acts in terms of a minimal frame, which includes only the directly specified consequences of the act and excludes other outcomes (e.g., the bettor considers only the likelihood of winning or losing the current bet). The minimal frame simplifies the evaluation process and reduces cognitive strain. However, occasionally a subject adopts a more inclusive frame (e.g., the bettor considers the likelihood of winning or losing the current bet in the context of having lost the previous one).

In general, researchers who design utility elicitation tasks intend to create health state descriptions to be interpreted with a minimal frame of reference. However, it is not possible to specify every aspect of a health state and the residual ambiguity may affect the frame of reference. As a consequence, subjects may adopt a more inclusive frame. Even when the health state description is extraordinarily detailed, subjects may adopt a more inclusive frame than the authors intended. Personal experiences with the health state being described may lead subjects to adopt more inclusive frames (i.e., during the elicitation subjects remember possibilities they might have overlooked initially; the memory may increase the salience of the health state and
hence its perceived likelihood) (Redelmeier et al., 1995). Therefore, simply specifying the health
description in greater detail does not necessarily lead to the adoption of a minimal frame of
reference, since the more that is specified, the more likely it is to induce the perception of
salience.

The objective of utility elicitation procedures is to obtain standard gamble utilities for use
in health-economic evaluations that accurately reflect respondents’ underlying preferences. The
main purpose of this paper is to develop and evaluate methods to detect inconsistencies in
standard gamble utility responses. Specifically, we discuss how respondents’ adoption of minimal
and inclusive frames affects the interpretation of inconsistency.

Example of Standard Gamble Utility Elicitation

In order to illustrate these consistency validation methods we use standard gamble utility
data from 101 healthy volunteers for health states defined by dependence in the single and
combined personal self-care capacities known as the activities of daily living (ADLs) (eating,
bathing, walking, dressing, continence, toileting, and transferring in and out of a bed or chair). The
utility elicitation is described in greater detail in the Appendix. Briefly, respondents were asked to
give 16 standard gamble utilities: first for their current health, then for each of the seven single
ADL dependencies (presented in a random order to each subject to balance order effects), then for
the health state of dependence in all 7 ADLs, and finally for 7 other combinations of ADL
dependencies (two combinations of two ADLs, two combinations of three ADLs, and one
combination each of 4, 5, and 6 ADLs). Of the potential 127 combinations of ADL dependencies
\(2^7 = 128\) minus the one possibility of not asking any of the possible combinations), only the 30
combinations found to account for 98% of ADL dependencies among the elderly were presented
to respondents (NLTC5, 1996).
Definition of Frame of Reference for Use in Inconsistency Assessments

We assessed utilities for inconsistencies first by assuming that the subjects had a minimal frame of reference and then by assuming that the subjects had an inclusionist frame of reference. If the health states of dependence in walking, bathing, and dressing are represented as in Figure 2, then we define subjects with a minimal frame as those who interpret the health state of dependence in walking and bathing by the area represented by region $\beta$ (i.e., dependence in walking and bathing but not dependence in dressing). We define subjects with an inclusionist frame as those who interpreted the same health state of dependence in walking and bathing as the areas represented by regions $\beta$ and $\gamma$ — that is, their frame included other ADL dependencies that were not specified. There are multiple possible “inclusionist perspectives” that a subject may apply to a health state. Our definition is of a particular inclusionist perspective, in which the respondent assumes that other ADL limitations may also be present when the program asks only about a subset.

We describe our methods for consistency assessments using the following set of hypothetical utilities for the health states of dependence in walking, bathing, and dressing:

<table>
<thead>
<tr>
<th>Health State: Dependence in the following ADL(s):</th>
<th>EU</th>
</tr>
</thead>
<tbody>
<tr>
<td>Walking (W)</td>
<td>0.94</td>
</tr>
<tr>
<td>Bathing (B)</td>
<td>0.93</td>
</tr>
<tr>
<td>Dressing (D)</td>
<td>0.92</td>
</tr>
<tr>
<td>Walking and Bathing (WB)</td>
<td>0.93</td>
</tr>
<tr>
<td>Walking and Dressing (WD)</td>
<td>-</td>
</tr>
<tr>
<td>Bathing and Dressing (BD)</td>
<td>-</td>
</tr>
<tr>
<td>Walking, Dressing, and Bathing (WDB)</td>
<td>0.88</td>
</tr>
</tbody>
</table>

The health states of dependence in WD and BD are left blank to simulate the scenario in which these combinations are not offered to this subject for rating.
Inconsistency Assessment Given a Minimal Frame of Reference

For those subjects with a minimal frame of reference, the following set of equations applies:

\[ EU_{W, \text{not-B}} = P_\alpha = 0.94 \]
\[ EU_{B, \text{not-W}} = P_\delta = 0.93 \]
\[ EU_{D, \text{not-W}} = P_\lambda = 0.92 \]
\[ EU_{W, B, \text{not-D}} = P_\beta = 0.93 \]
\[ EU_{W, D, \text{not-B}} = P_\varepsilon \]
\[ EU_{B, D, \text{not-W}} = P_\kappa \]
\[ EU_{W, B, D} = P_\gamma = 0.88 \]

where \( P_i \) is the probability defined by region \( i \) in Figure 1 and \( EU_j \) is the reported utility for dependence in ADL\(_j\). If we assume that subjects consistently maintained a minimal frame, there is little that can be said about consistency on the basis of joint probabilities. Each probability region would be non-overlapping (\( i.e. \), they would be disjoint probabilities) and we would have to conclude from a perspective of joint probabilities that all responses are consistent. If we wanted to impose a definition of consistency that included ordering of utilities such that single ADL dependencies are preferred to combinations of dependencies (we call this intuitive ordering), we could not do so on the basis of a probabilistic assessment because the regions are disjoint.

Inconsistency Assessment Given an Inclusionist Frame of Reference
For those subjects with a maximally inclusive frame, the following set of equations would describe the relationship between reported utilities and the joint probability regions associated with each health state:
\[
EU_W = P_a + P_b + P_\gamma + P_e = 0.94 \\
EU_B = P_b + P_\delta + P_\gamma + P_\kappa = 0.93 \\
EU_D = P_\gamma + P_e + P_\kappa + P_\lambda = 0.92 \\
[1] EU_{WB} = P_\beta + P_\gamma = 0.93 \\
EU_{WD} = P_\gamma + P_e = \\
EU_{BD} = P_\gamma + P_\kappa = \\
EU_{WBD} = P_\gamma = 0.88
\]

Because these regions are joint probabilities, we can determine consistency by evaluating the extent to which the reported utilities conform to the axioms of probability theory (i.e., determine whether or not the reported utilities can arise from some probability distribution). A given subject will have provided utilities for some but not all of these states (because each subject provides utilities for 15 different combinations of ADL dependencies). We can therefore determine whether a nonnegative solution to this set of equations exists. Moreover by applying a variety of constraints to such a solution, we can assess the degree of inconsistency.

Perhaps the most intuitive and least subtle measure of consistency is that utilities should be ordered such that utilities for combinations of ADL dependencies should be less than or equal to subsets of that combination of dependencies. In the case of the example this means that we expect \( EU_{WB} \leq \min (EU_W, EU_B) \), \( EU_{BD} \leq \min (EU_B, EU_D) \), and \( EU_{WBD} \leq \min (EU_{WB}, EU_D) \). A subject whose responses conform to these inequalities would be responding in a manner such that his utility for dependence in walking and bathing was consistent with his utility for dependence in walking alone. We call this "order constrained consistency". Note that in "intuitive ordering" the subject is assumed to have a minimal frame and that in "order constrained consistency" the subject is assumed to have an inclusive frame of reference.

We can extend the constraints applied to a set of utilities to evaluate more subtle inconsistencies than simple misorderings of utilities. Consider, for example, a set of utilities:
$EU_{W} = 0.94, EU_{B} = 0.93,$ and $EU_{WB} = 0.85$. These utilities are order constrained consistent. However, $EU_{W, not-B} = 0.09, EU_{B, not-W} = 0.08,$ and hence $EU_{W, B} + EU_{W, not-B} + EU_{B, not-W} = 1.02$, which exceeds 1.00. When order constrained consistency occurs and the sum of the constituent probabilities are less than or equal to 1.00, then we say that the responses are "moderately constrained consistent".

Finally, we can constrain the solution further by adding the equation $P_{\alpha} + P_{b} + P_{S} + P_{Y} + P_{t} + P_{\lambda} + P_{k} + P_{\mu} \leq 1$ to the set defined in [1]. When the sum of the constituent probabilities is less than or equal to 1.00, we say that the system is "strongly constrained consistent". We wrote a program of linear equations in nonnegative variables to perform these consistency validations (Bravata, Cottle, Eaves, & Olkin, In Press). For each subject, we first assumed that he had a minimalist frame and asked if his utilities met the definition of intuitive ordering consistency. We then assumed that he had an inclusionist frame and determined whether his utilities met the definitions of order constrained, moderately constrained, and strongly constrained consistency. Additionally, under the assumption of the inclusionist frame, we calculated a measure of the degree of inconsistency and the number of ADLs in the combination when the subject was first noted to be inconsistent. This measure of degree of inconsistency is not on an easily interpretable scale (i.e., whereas a value of 0.01 represents less inconsistency than a value of 0.1, it is not 10 times less inconsistent); however, it provides a measure for distinguishing between the most and least consistent subjects and can be used as a threshold of inconsistency for subsequent analyses of data collected on utilities.

**Inconsistency Assessments**

Assuming a minimal frame for all 101 subjects, 39 were inconsistent by applying an intuitive ordering such that dependence in a subset of any combination of ADLs was less than or equal to dependence in the combination. Assuming an inclusionist frame for all 101 subjects, 39 were inconsistent by order constrained, 58 by moderately constrained, and 73 by strongly constrained solutions. The 39 subjects identified as being inconsistent by intuitive ordering and the order
constrained solutions were the same individuals — that is, they were inconsistent regardless of frame. It is important to note the relationship in order consistency between approaches assuming a minimalist versus inclusionist frame. For example, if we wanted to know if a subject had ordered $EU_{w,B} \leq EU_{w,\text{not-B}}$, for the minimalist frame we would ask: Is $P_\beta \leq P_\alpha$? Whereas, for the inclusionist frame, we would ask: Is $P_\beta + P_\gamma \leq P_\alpha + P_\beta + P_\gamma + P_e$? If the latter equation is true, the former is also true.

Among the inconsistent subjects, most were inconsistent in only a few of their responses. By the order constrained solution, 62% of the subjects had 4 or fewer inconsistencies (Figure 2). Similarly, by the strongly constrained solution 66% had a measure of degree of inconsistency less than or equal to 0.1 (Figure 3).

Inconsistency with increasing ADL dependencies. During the elicitation process, subjects were first asked to give utilities for dependence in single ADLs, then for combinations of two ADLs, then combinations of three and so on. In order to determine whether the frequency of inconsistencies increased with the number of ADLs in the combination, we plotted the proportion of inconsistent subjects identified by each of the three solutions versus the number of ADLs in the combination when the subject was first identified as being inconsistent (Figure 4). The strongly constrained solution detected 78% of inconsistent subjects when they gave their utilities for dependence in combinations of 2 ADLs. These subjects were the same individuals who were identified as inconsistent by the order constrained solution albeit only after providing utilities for higher order combinations.

In this small data set, the consistency reminder was not significantly associated with enhanced consistency. Given a sample size of 101, of whom 38 received the consistency reminder, we lacked sufficient statistical power to detect small differences in the proportion of inconsistent subjects between the group who received the reminder and the group that did not.

Differences in utilities between consistent and inconsistent subjects. Mean utilities for the 39 inconsistent subjects detected by the ordering constraining solution were compared to the 62 consistent subjects. The utilities for dependence in continence and the combination of all 7
ADLs were significantly higher among the consistent subjects than inconsistent subjects ($p \leq 0.002$).

**DISCUSSION**

We developed a model for evaluating variations in consistency on the basis of framing effects and tested a novel method of consistency validation that can be applied to utilities of multiattribute health states. We found that potential inconsistencies in multiattribute utility data occur frequently and that the utilities of consistent subjects are significantly higher than those of inconsistent subjects. We found that the estimated degree of misordering does not depend on whether subjects were assumed to adopt a minimalist or inclusionist perspective.

It is important to note that these are potential inconsistencies and that “inconsistent” subjects might be able to explain how their utilities reflect their underlying preferences. Preference reversals, or other so-called errors of choice or judgment, are not necessarily irrational. The practice of acting on the most readily available frame can sometimes be justified by reference to the mental effort required to explore alternative frames and avoid potential inconsistencies (Kahneman & Tversky, 1981). Additionally, utility data are most often collected in the absence of an accompanying explanation from the subject; it is impossible to know which of the utilities that seem to violate researchers’ expectations reflect underlying preferences and which result from poor task comprehension and shifts in frame. In the absence of such an explanation, it is unclear which of the apparently inconsistent utilities should be used in the analysis and which should be discarded. The choice of which to include could have significant effects on reported utilities as they did in our study.

Our linear programming (LP) method can be used to evaluate the consistency of utilities. It is likely to be particularly useful for large sets of multiattribute utilities, when it would be difficult and time-consuming to detect potential inconsistencies manually. It can be used to generate several measures of inconsistency that may be useful for evaluating the mean utilities of a sample population. The degree of inconsistency could be used to set thresholds for including
subjects’ utilities in the calculation of the population’s mean utility. The strongly constrained LP solution can detect potential inconsistencies early in the elicitation, which suggests that a computer-based elicitation program coupled with a linear program could detect potential inconsistencies during the elicitation process and provide an opportunity for a consistency enhancing intervention. The advantages of our LP approach include its strong foundations in probability theory, its ability to detect very subtle inconsistencies, and its flexibility for application to complicated multiattribute elicitations. Its principal disadvantage is its underlying assumption of an inclusionist frame. Whereas some subjects may have shifted to a more inclusionist frame of reference during the elicitation process, especially for the complex sets of five and six ADL dependencies, it is not possible to determine the extent to which such a shift occurred. Additionally, whereas subjects should be able to state their preference for ordering of health states, it is highly unlikely that any subject could calculate the probability distributions required for consistency by the linear program.

Despite these limitations, our study has important implications for multiattribute utility assessment. First, even though subjects were instructed to provide a relatively large number of utilities (n = 15), it does not appear that inconsistencies resulted from fatigue. Given that the strongly constrained solution detected 70% of inconsistent subjects when they provided their utility for the first pair of ADLs presented to them (early in the elicitation), it seems unlikely that the inconsistencies found in our population were due primarily to fatigue (Figure 4). Second, the differences in utilities between the consistent and inconsistent subjects are not readily explicable. It is unclear why the two groups should differ in their utilities for continence and all 7 ADLs and not the other ADLs.

The designers of the software used to elicit the utilities for this analysis gave extensive attention to creating an elicitation device that would obtain the most accurate estimates of subjects’ true preferences. To minimize framing effects, all risks were shown as both risk of death and chance of perfect health. The health state descriptions were designed to present an unambiguous concept of dependency in an ADL in a balanced way that did not interject
assumptions about emotional states. The program allowed respondents to review previously described health states and to change their preference ratings after further consideration. The review screens allowed subjects an opportunity to revisit and change their reported utilities. In the first review screen, subjects had an opportunity to re-evaluate their ratings for each individual ADL after having considered each of the 7 individually and also having considered the health state of dependency in all 7. Similarly, the second review screen encouraged re-evaluation of ratings for combinations of 2 to 6 ADL dependencies after having considered the progressively more disabled states of increasing numbers of ADL dependencies. The developers included these review screens in order to enhance accuracy by maximizing the likelihood that subjects would distinguish one ADL from another and also would fully appreciate dependencies in multiples ADLs as a possibility when considering the utility of dependence in a single ADL. Nevertheless, the cognitive complexity of comparing, for example, one set of 4 ADLs with a set of 5 somewhat different dependencies may be too challenging for most subjects to manage without a shift to the inclusive frame.

Given the high prevalence of inconsistency and the finding that utilities of consistent and inconsistent subjects differ, we recommend that utility assessments for multiattribute health states include efforts to measure and minimize inconsistencies. The LP approach we describe is one method for doing so. Additionally, we recommend additional research that explores why subjects give what appear to be inconsistent or apparently invalid ratings. This work is critical for verifying that the utilities intended for use in cost-effectiveness analyses represent the best estimates of subjects’ preferences.

ACKNOWLEDGMENTS

We thank Sheryl Davies, Jennifer Zeigler, and Manijeh Parineh for their assistance with the subjects and data collection, and Lyn Dupré for her review of the manuscript.
FIGURE LEGENDS

Figure 1 Graphic representation of the standard gamble.

Figure 2 Venn diagram of joint probability regions for the example of dependence in three activities of daily living. This Venn diagram depicts the joint probability regions for dependence in walking (W), bathing (B) and dressing (D).

Figure 3 Order constrained inconsistencies per subject. Inconsistencies per subject are plotted along the x-axis and cumulative percent of inconsistent subjects along the y-axis. The total number of order constrained inconsistencies are presented in the upper panel. The measure of degree of inconsistencies is presented in the lower panel with the moderately constrained data by the small-dash line and the strongly constrained data denoted by the large-dash line.

Figure 4 Number of ADLs in the combination when inconsistency is first detected. The cumulative proportion of inconsistent subjects is plotted on the y-axis and the number of ADLs in the combination when the subject was first identified as being inconsistent along the x-axis. The order constrained data are denoted by the unbroken line and the strongly constrained data by the large-dash line.
Figure 1

Alternative 1

- Probability = $p$ (Best State, Utility = 1)
- Probability = $1 - p$ (Worst State, Utility = 0)

Expected Utility = $p$

Alternative 2

- Probability = 1 (State Being Rated, Utility = $p$)

Expected Utility = $p$
Figure 3

Cumulative percent of inconsistent subjects

Total number of order constrained inconsistencies per subject

- Moderately constrained
- Strongly constrained
Figure 4
REFERENCES


APPENDIX. UTILITY ELICITATION DESCRIPTION

The computer presentation began with an introduction that included a general introduction to the pointer, trackball, and the buttons on the screen. Respondents were then offered an opportunity to practice using the trackball and clicking the mouse. They observed an explanation of self-rating of dependency and descriptions of each ADL dependency. They then practiced giving standard gamble utilities for two examples (having a head cold and being blind). Finally, respondents were asked to give 16 standard gamble utilities: first for their current health, then for each of the seven single ADL dependencies (presented in a random order to each subject to balance order effects), then for the health state of dependence in all 7 ADLs, and finally for 7 other combinations of ADL dependencies (two combinations of two ADLs, two combinations of three ADLs, and one combination each of 4, 5, and 6 ADLs).

When asked to provide their utilities for the single ADL dependencies, the subjects were presented with the following text:

"In this program we are interested in finding out how you feel about the following activities: bathing, continence, dressing, eating, toileting, transferring, and walking. We have described what it is like to always need daily help with these activities. Now we will ask you to imagine having a health condition that interferes with your ability to perform these activities, if you do not have such a condition already. Having the condition means that to perform this activity, such as eating or dressing, you need daily help. You would NOT need help in any of the other activities we have previously described. We will ask you to consider the risks and costs you are willing to bear for a treatment that would improve your health so you would no longer need any help with these activities. The technique we will use is called the ‘Percent Method’. This technique poses a simple question: Imagine there is a treatment that would completely cure you of ever needing help with a particular activity, such as dressing yourself. But, this treatment has risks. In fact, this
treatment poses a risk of death. That is, if you undergo this treatment for a complete
cure, there is a chance that you will die. What is the highest chance of death that you
would risk for a treatment that would completely cure you?"

The risk of death that respondents were willing to accept to avoid dependence in a given
ADL or combination of ADLs was obtained as follows. Subjects were asked if they would be
willing to accept a 1% risk of death and 99% chance of cure. If they answered yes they would be
asked the question again. This time they would be presented with probability values that varied
in a converging ping-pong fashion, that is, by alternating between high and low values: (90% risk
of death, 10% chance of cure), (10% risk of death, 90% chance of cure), (80% risk of death, 20%
chance of cure), and so on. After answering the first two yes/no questions, subjects could directly
provide the probability of death at which they were indifferent between the health state of ADL
dependence occurring with certainty and the gamble by clicking on up/down arrows to change
the numerical value presented to them.

Subjects were given two opportunities to review their ratings and to make changes, first
after rating the health states of dependence in single ADLs and dependence in all 7 ADLs and
again after rating the combinations of ADL dependencies. The review screens showed all the
ratings given in the previous segment and invited the subjects to review them and make any
changes they wished. Additionally, 38 subjects were randomly selected to receive a computer-
generated consistency reminder. If one of these subjects rated any health state worse than the
combination of all 7 ADLs, he would receive an alert, showing his rating for dependence in all 7
ADLs, and calling attention to the lower rating he had given to a single or combined ADL
dependence. He was then given an opportunity to reconsider his response. This reminder was
presented only once for any inconsistent rating; that is, it did not force the subject to change his
response.