Utility Assessment Based on Individualized Patient Perspectives

Duane Steward, DVM, MSIE
Clinical Decision Making Group; Laboratory for Computer Science, M.I.T.
NE43-415; 545 Technology Square
Cambridge, MA 02139
duane@mit.edu

Abstract

A novel approach to utility assessment is presented which enables the elicitation of an individualized understanding of the patients' perspective. The result illuminates what dimensions of health are regarded as important in the patient's own terms. This approach is in contrast to traditional utility assessment, which is conducted in prescribed terms, is difficult to apply with confidence on an individual basis and solicits a non-reusable single value for a specific outcome of a decision context. This paper discusses a method which attempts to elicit a complete, comprehensive and operational system of attribute scales to describe the relevant attributes of health in an individual's personal perspective. The result is a personalized model that can be used to score any health in those terms with re-usable components. Piloting investigations with this fundamental approach reveal potential as a foundation for simple or complex multiattribute modeling and as promising with respect to psychometric evaluation. The systematic approach is grounds for expectation of automation and consistency which could raise the level of applied decision analysis in medical decisions.

Motivation and Background

Decision analysis is an elegant technology for explicating the issues of medical decisions which involve tradeoffs. It is difficult, however, to apply this technology to individual patients' decisions when the utility value for each outcome is unclear. This is most often the case because of the diversity, uncertainty and ambiguity of patients' values. It is made worse by a provider's use of unfamiliar terms and absence of a systematic approach to understanding the individual patient's way of viewing the specific medical decision. Consequently, despite a two decade history of pioneering application to medical problems, decision analysis is absent in common medical practice. In an effort to ameliorate these problems, this paper describes a novel approach to capturing patient utility values using attributes for healthiness that are elicited directly from the individual patient.

Von Neuman- and Morganstern-based utility theory (Von Neuman and Morganstern 1976) provides a coherent basis for elicitation methods which, with maturing health status measures, are enabling a growing impact on health care policy and resource management from a social and corporate perspective. One glaring absence is the means to represent the individual's perspective if it differs significantly from the empirical norm of studied or consulted populations. Traditionally, health status measures have reflected the value system of the biomedical community (e.g., life span was used long before quality of life adjustments were included). Recent recognition of this biased view of life has given rise to the question of what dimensions should be included in health utility assessment. Most recently this has been addressed through the efforts to elicit the most meaningful parameters from experts and patients gathered in "focus groups" intending to develop a model for what dimensions to include. However, as a matter of practicality, dimensions rarely shared by more than a few persons are neglected in this approach. This does nothing to accommodate those individuals for which these dimensions serve an unusual importance. It serves the purpose of society but not the individual.

Fundamentally, the practice of utility assessment is not incapable of individual application. The commonly recognized methods (rating scale, standard gamble and time trade-off) have been applied to a lengthy list of medical domains, one patient at a time. In the absence of a satisfying "gold standard", ambivalence about the results exists on the part of the health care provider. This and patient violation of prescriptive conclusions have prompted modified methods. Repeatedly in this evolutionary process, the evidence describes diversity among individuals and contexts, which so far eludes a coherent overall explanation outside the assertion that people are often irrational. The complexity of instruments and results underscore the demand for trained experts to conduct utility assessment where it is required. Skill required to maintain quality control on the elicitation process is exceeded by that required for interpretation of the answers given by those interviewed, as well as that required to stay abreast of the growing list of questions raised about the methodology and their ramifications. Not only must the analyst be highly trained, they must also slow down and take time to actively listen to the often inarticulate person from whom the information must be acquired. Rushing it confuses and confounds.

For current methodologies to acquire meaningful information, utility assessment must force the person responding to formulate answers to grave potential outcomes for their life. If the technology is constrained to

application in demanding circumstances, the individual is likely faced with very personal encounters with their mortality which usually results in severe emotional stress before the elicitation is complete. Not only is this emotionally painful but it is embarrassing for both the patient and the analyst, further discouraging the use of such technology except in extreme cases.

Regarding individualized application, the result is a utility assessment procedure that is expensive and requires the skills of a highly trained expert which will not be conveniently available. Specifically, this means a member of the staff of the Clinical Decision Making Division of the New England Medical Center famous for implementing decision analytic technology can only attest to performing utility assessment on an individual patient a few dozen times in half as many years (Eckman 1997). When performed, the process takes two hours and usually leaves the patient in tears. The assessment must be performed during regular hours of the professional day most likely in the sterile environment of a referral health care institution.

The result of the individualized utility assessment, even when a combination of rating scale, standard gamble and time trade-off methods are employed, is no more than a gross proxy measure. For all the effort, the single value for each health state evaluated is an overall summary measure for everything involved in that specific health state, predicating assumptions aside. Should the context of the decision change, i.e., the potential outcomes shift in their description or context, the procedure must be repeated from scratch. Should the patient change their outlook on life, a goal for much of medical therapy and provider patient interaction, the results of such former utility assessments are equally obsolete. No parts of the all encompassing proxy measure are reusable.

There is a great need for a systematic approach to the discovery of what is and what is not important to the individual and an operational understanding of the patients' perspectives if their preferences are to be used in medical decisions.

In an effort to address these issues and the uncertainty of current utility assessment results, I have sought a methodology which would better understand a person presumed rational and expert on their own view of healthiness. This method would first be descriptive and then operational in that the ability to communicate the values of the patient would enable a caring provider and patient to collaborate in health care decisions using values expressed in the patients terms. A systematic approach would reward both parties with a sound way to expose deficiencies in either's perspective and the opportunity to explicitly address such details rather than proceed on ignorance or assumption. Such an approach would explicate a complete list of attributes with which an patient builds his or her view of healthiness. Each attribute contributing should be comprehensive. A model for composing the complete list of comprehensive contributing attributes would provide a similar value for a specified health state to that provided by traditional summary measures. A decomposed multiattribute approach, however, would suggest possible explanation for unanticipated summary scores as well as potentiate reuse of robust component parts. Upon arbitration and resolution the resulting value system could be employed to express the decisions faced in the familiar terms of the individual and, based upon the facilitated understanding of the individual, result in improved success attaining mutually prescribed goals.

In pursuit of such a methodology, I propose a method which elicits a person's individual list of dimensions of healthiness on the basis of self disclosed means of discrimination between friends and family. The identification of a complete list of dimensions results in what will be referred to as an ordered attribute scale for each dimension (or attribute) of relevance to the individual. Each scale should be comprehensive, i.e., cover the extremes possible (in the individual's perspective or experience) for that attribute. Each scale should be operational on the basis of the semantic referral to a measurable construct in the hands of those other than the patient and on the basis of understanding the utility of intermediate attribute levels relative to the extremes which could be normalized (e.g., at 0 for worst and 1 for best). The second phase of elicitation, then, is to identify the numeric position of each intermediate attribute level relative to the extremes. Finally, a third phase, elicits the necessary coefficients for a composing model which enables the user of the output to ascribe multiattribute scores and summary measures to any health state for which the level of each of the attributes can be ascribed. For the purposes of development I begin with a simplifying assumption of a linear additive model for composition. Multiplicative and more complex models could be employed as will be mentioned in the discussion, but are beyond the scope of this paper.

As will be seen, the procedure does not involve an explicit health state description and thus is not context dependent. It only relies upon the modeling assumptions and the stability of the individuals perspective on life. The systematic nature potentiates the use of programmed elicitation which free the process from timing and location constraints mentioned above. This freedom suggests more frequent opportunity to perform assessment on the order of frequency that a person might change their perspective on health and their preferred basis for medical decisions. In addition, the freedom from specific health contexts renders the value system contained as generic and reusable for unanticipated outcome descriptions of an evolving decision tree model.

Values Elicitation Method

Determination of Attributes

The proposed method for eliciting utility values begins with the determination of attributes by which the patient discriminates healthy friends and relatives from unhealthy friends and relatives. These attributes must be measurable to support the overall methodology objective (see discussion of operational attributes in the background section). An assumption upon which the methodology depends is that all immeasurable attributes can be decomposed into measurable 'sub-attributes' if pursued. Once an attribute is determined by naming the extremes, the task remains to determine the scale by which the attribute may be measured. An ordered set of two or more categories is called an "attribute scale" with extreme and intermediate levels. Once the labels are determined for each possible level an attribute might have, a mapping function to utility values (standardized to range between 0 and 1) must be determined. For dichotomous scales, this is trivial in that the true or positive value of the binary scale maps to the 1.0 extreme and the false or negative value to the 0.0 extreme of the utility function. The ordered nominal scales, however, can be acquired in two steps. First, the elicitation of terms to label all possible intermediate values of the attributes (assuming the extremes of the attributes are what have been acquired to begin with). Second, the appropriate utility value for each of those attribute values can be determined by employing standard utility assessment (e.g., gamble or trade-off methods comparing the intermediate attribute value to a lottery between the extreme attribute values or to the optimal extreme respectively). The result thus far is a look-up table or function (for continuous interval scales) for all potential values that might be used to score health states, however, thus far no specific health state is required for elicitation.

Rules regarding the formulation of subsequent questions provide structure to the interview. These rules are designed to preserve the absence of bias while focusing the conversation on the desired elicitation. An informal presentation of those rules follow in a description of the interview details. Graphic visual aids facilitate the understanding of the questions throughout the interview. The first pair of questions is to be the same for all participants. Each subsequent question is formulated from the previous responses.

The interview begins with asking the subject to think of three friends or family members specified by a role or relationship, e.g.,- father, best-friend, best-friend's father. Three roles are presented for the subject to fill with someone they know. Names are not required but the age of the role-filler is recorded for clarity and to facilitate any necessary reference during subsequent discussion. Once this introduction is completed the first question follows. That question is, "With those three persons in mind, tell me an important way in which two of them are alike as more healthy than the third. If it is easier, you may choose to tell me how two are alike as less healthy than the third." Each separate response is regarded as one end of an attribute scale. It represents one dimension by which the patient discriminates healthiness from unhealthiness. For each such response, the person is asked, "How, then, is the third less healthy [or healthier] than the other two?" This provides the 'first draft' of a bipolar attribute. The same triad of roles may be employed for multiple bipolar attributes or changed if needed to solicit fresh ideas. The interview continues until no more ideas come to mind. The patient is encouraged with this question to think of any attributes not represented in the specific people brought up in the interview. One of four sets of roles are used to focus the subject on three familiar people of whom they may have enough acquaintance so as to hold some opinion of their healthiness. Convenient substitution is allowed. The specific role filled is not so important as that three distinct individuals of satisfactory familiarity are fixed in mind. Example sets of roles used are as follows:

Your father or nearest male relative with whom you have most frequent contact
Your best male friend
Your best male friend's father

Your mother or nearest female relative with whom you have most frequent contact
Your best female friend
Your best female friend's mother

Your boss or employment supervisor
Your most frequent customer or client
Your friendliest business associate

Your nearest neighbor in any direction from your home (someone you are familiar with)
Your nearest neighbor in the opposite direction (again, someone you are familiar with)
Your most frequent neighborhood visitor who does not live nearby

Additional properties of the attribute scale are solicited to verify understanding and for purposes of validating comparability with other responses from the same person or other interviews. Providers interested in changing the patient's behavioral risk factors will have an interest in certain properties that these bipolar attributes possess. These properties might include activity, origin, control, causality, implication and/or measurability. These properties may be solicited with the following sorts of questions. Is the concept of the attribute scale one that requires energy or happens without energy required (activity)? Is the nature of the attribute scale something

which comes from within a person or something that comes from somewhere outside the individual(origin)? Is the concept of the attribute scale something which in within the individual's control or something that is beyond the control of the individual (control)? Does the concept of the attribute scale cause healthiness/unhealthiness or is it a result of being healthy/unhealthy; both or neither (causality)? Measurability is determined by the interview administrator or self reporting such that if the response is deemed unmeasurable, the patient is asked for explanatory subordinate concepts with a question framed from the response, e.g.,- "What kind of a person is one who is indicated healthy by [response substituted]?" To minimize the length of the interview, past experience with interviews could be employed in implementations that follow the first prototypes. If the response is like one heard previously within the same interview or outside interviews, the interview may simply ask for verification of the properties as registered in previous experience as a less taxing elicitation.

It is worth noting that the chosen wording of the questions asks for the means by which the person discriminates healthy from unhealthy friends and family. This choice is not arbitrary, rather it capitalizes upon the experience of those who have developed the repertory grid elicitation process (Fransella and Bannister 1977) (Ryle 1975). Questions are carefully fabricated to avoid the ambiguity of feelings about healthiness. They instead focus on the characteristics by which persons discriminate healthy from unhealthy. Before the interview is completed, they will be asked to evaluate their willingness to trade one state of health for another; each described solely in these personal terms. If the dimensions prove inadequate for such an exercise, the opportunity to revise remains.

Each session is ended with a final query: "Now that you know the sort of thing we are looking for, are there any other ways in which you discriminate between healthy and unhealthy people regardless of any particular person we have brought to mind?" The patient is encouraged with this question to think of any attributes not represented in the specific people brought up in the interview.

The informal rules of the interview are then:

1. Refer to the "healthy" or "unhealthy" root attribute as an anchor for each question.
2. Each new elicitation begins with "How are two alike as more[less] healthy than the third?"
3. Each item elicited must be paired with a contrasting partner term by asking, "...how then is the third less[more] healthy than the other two?"
4. End the query when the subject cannot think of any more means by which they might discriminate the individuals of the role triad based upon their healthiness.
5. The subject will be asked to express their personal standing in regard to each elicited attribute as a demonstration that the interviewer properly understands the bipolar attribute as they intended it (discussed below).

Determination of intermediate levels of attribute scales

The next step of the values elicitation process is to solicit intermediate levels the attribute scale may potentially have, if any. The interviewed patient is asked if they can think of any person who does not belong to either extreme of the scale but somewhere in between. If not, the assumption is that no intermediate levels exist in the perspective of the patient. If a person is brought to mind, specifically or categorically, the patient is asked for a categorical label for such a person. This query is repeated for each bisection created until possibilities are exhausted. The result is an ordered series of nominal categories which are ordered by nature and constitute all plausible levels of the attribute in the patient's perspective. The patient can be asked directly if each extreme, in fact, covers the most extreme cases that can be imagined. Attribute scales from previous interviews may be less taxing to verify or modify than built from scratch and therefore future developments of the methodology should include some utilization of past responses without undue bias.

To further validate the understanding, the interview involves formulating a question from each attribute scale elicited. The purpose of this step is to test the interviewer's understanding. If we can agreeably phrase a question with their bipolar attribute, higher confidence is associated with the understanding. The question to be formulated is one which asks the subject to rate their own personal standing on the attribute scale. "You seem to indicate that a person who is <first elicited pole> is healthy and one who is <contrast pole> is not so healthy. Do you see yourself as healthy because you < first elicited pole> or as unhealthy because you are <contrast pole>?" For scales with more than two levels, the question should be framed to include all levels as potential responses. Explanation is to be offered to the subject of the fact that the answer does not matter as much as the question. It is simply an opportunity for the subject to recognize any misunderstanding and provide opportunity for restatement or clarification prior to approval.

The attribute scale for which all levels plausible are known, is then assessed to determined the utility values for each level. Utility assessment techniques of Keeney and Raiffa (Keeney and Raiffa 1976) are employed to determine utility values for attribute scales with inherent cardinality. The ordered attribute scales left have two or more discrete levels. For scales with only two levels, the extremes of 0 and 1 are assumed for the worst and best ends of the attribute scale respectively and the elicitation of the attribute scale is complete. For scales with more than two levels, the interview proceeds with a modification of

the standard gamble utility assessment. The patient is told to hypothetically assume they exhibit the level of attribute being assessed and offered a hypothetical potion which promises to change them to the best possible level of the attribute but at some risk of ending up in the worst level of the attribute. The specific risk is systematically varied and as in the earlier description, the probability at which the patient is indifferent or cannot decide between options is the utility value assigned to that attribute level. Similar potential exists for employing a similarly modified rating scale or time trade-off method to determine the utility value of intermediate levels relative to the extremes of the dimension of health represented in each attribute scale. In any case, the result is a mapping function which if given the name of the level for any of the individual's attribute scales, provides a numeric value between 0 and 1 for that attribute of health.

Determination of combining model for health state scoring.

To determine the relative weights of each attribute scale for a composite scoring of health states under the linear model assumptions, the patient is asked the following series of questions. First they are asked to assume that they are currently in a hypothetical health state described completely by one level from each of their elicited attribute scales. Then they are asked to assume they have a life expectancy of a given number of years in that state, e.g., 10 years. They are asked if they would trade that outcome for some number of years in a different health state described in the same manner, e.g., 5 years in a state with a higher level of one attribute. The number of years in the second state is varied until the point is discovered where the preference changes indicating equal value for the two hypothetical states of health endured for different amounts of time. Similar equivalencies are sought until a differential in time is scored for a change in each attribute scale granting the circumstance where the coefficients may be computed by solving for the simultaneous equations.

To illustrate the method, imagine the oversimplification where a patient measures health by only three two level attribute scales: exercise (some or none), dietary restrictions (none or any), and presence of any disease (true or false). The patient is asked to imagine they will have a life expectancy of 10 years as an individual whose state is best described as S1={renal disease, protein restricted diet, and no exercise }. Further they are asked if they would trade that state of health for one with 5 years life expectancy described as S2={no disease, protein restricted diet, and exercise }. For illustration, assume it is determined that the patient is ambivalent about the choice between 10 years in the first (S1) and 7 years in the second (S2) state of health. The patient is also asked similarly for the number of years they would trade for in a third state S3={no disease, no dietary restrictions, and no exercise}; let's say it is 9 years. With the values known for the levels of each attribute (1.0 or 0.0 in this case) only the scaling coefficients of the linear model are unknown. Using the method of simultaneous equations we can solve for each scaling coefficient knowing that the sum of the coefficients themselves must sum to one.

$$\text{Utility}(S1) = \lambda_1\text{Attribute}_1 + \lambda_2\text{Attribute}_2 + \lambda_3\text{Attribute}_3 = \lambda_1(0.0) + \lambda_2(0.0) + \lambda_3(1.0)$$

$$\text{Utility}(S2) = \lambda_1\text{Attribute}_1 + \lambda_2\text{Attribute}_2 + \lambda_3\text{Attribute}_3 = \lambda_1(1.0) + \lambda_2(0.0) + \lambda_3(1.0)$$

$$\text{Utility}(S3) = \lambda_1\text{Attribute}_1 + \lambda_2\text{Attribute}_2 + \lambda_3\text{Attribute}_3 = \lambda_1(0.0) + \lambda_2(1.0) + \lambda_3(0.0)$$

and

$$\text{Utility}(S1)\ 10\text{yr} = \text{Utility}(S2)\ 7\text{yr} = \text{Utility}(S3)\ 9\text{yr}$$

therefore

$$[\lambda_1(0.0) + \lambda_2(0.0) + \lambda_3(1.0)]10\text{yr} = [\lambda_1(1.0) + \lambda_2(0.0) + \lambda_3(1.0)]7\text{yr} = [\lambda_1(0.0) + \lambda_2(1.0) + \lambda_3(0.0)]9\text{yr}$$

means

$$\lambda_1 = 0.587,\ \lambda_2 = 0.217,\ \textit{and}\ \ \lambda_3 = 0.196.$$

This means that the expected utility for this patient for any given health state is given by the formula:

Utility (given state of health) = 0.587 Disease Attribute Level + 0.217 Dietary Attribute Level + 0.196 Exercise Attribute Level

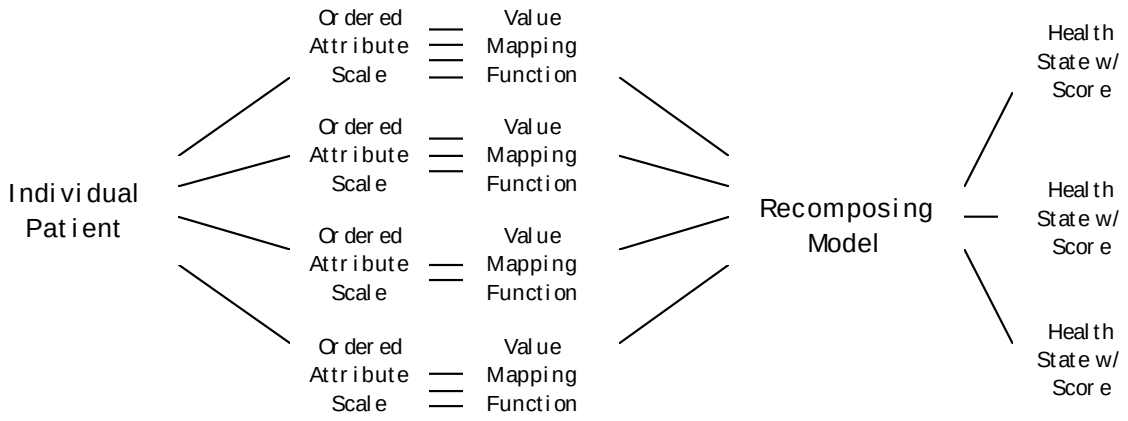

Figure 1. Overview of decomposition and re-composition of patient values. The individual's perspective of healthiness is composed of one or more attribute scales, each scale having a number of possible levels. Each level has a corresponding value on a numeric scale from 0 to 1. Knowing the level for a given health state for each attribute scale and the model for recombining the attribute scale scores in a composite score allows the use of the individual's perspective to score that health state.

Scale ID	Ordered Attribute Scales	Value Relative To Attribute Extremes
i	no chronic illness or disability	1
	temporarily ill	0.999999999
	minor health concerns	0.999999
	probability of becoming chronically ill but not yet	0.3
	chronic illness **	0
ii	active life	1
	hypochondriac	.53-.54
	somewhat active or limited	0.4
	forced sedentary life **	0
iii	mental well being **	1
	mentally ill	0.25
	mental decay-marked decline in level of mental functioning	0
iv	non-obsessive, Non-abusive **	1
	occasional lapses: into heavy drinking, fluctuating weight, yoyo dieter, binge drinking, recreational drug use(no addiction)	0.95
	obsessive personality, drugs or alcohol, self destructive	0

Figure 2. Ordered attribute scales of one individual interviewed. All levels of each scale are listed with their corresponding value relative to the scale extremes to the right. The four attribute scales for patient #3 are arbitrarily numbered on the left. The level which this individual claims best describes his current state of health is marked with two asterisks, "**".

Preliminary Results

The methodology presented here is the result of a parallel pursuit of human mediated interviews and codification of

the mechanics based upon those experiences. Interviews on the Boston subway which went no further than the elicitation of bipolar constructs (i.e., no elicitation of intermediate levels, mapping function values or relative weights) revealed that the quantity of attributes relevant for the tested population ranged from 3 to 25 with most cases falling between 5 and 15. The questions used proved to be feasible in that all volunteers were able to provide answers. The absence of refused interviews and enthusiastic well wishing for the research on the part of those interviewed was encouraging.

Interviews applying the entire protocol are now being conducted. Subjects interviewed are visitors to the Emergency Department of the Beth Israel Deaconess Hospital of Boston. Personally conducted by this investigator over the telephone in the weeks that follow an emergency visit, all interview volunteers have been able

Patient	Attribute Scales	Coef	HS0	HS1	HS2	HS3	HS4	HS5
1								
	Activity	0.3378	1	1	0.5	0	0	0
	Allergies	0.3344	1	na	na	na	na	na
	Get sick a lot?	0.3277	1	1	1	1	0	0
	composite utility value		1	1	0.831	0.662	0.334	0.334
	SG			70-75	65-75	50-52	45	35-40
2								
	Peace of Mind	0.1887	1	0.75	0.75	0.75	0.75	0
	Alcoholism	0.151	1	na	na	na	na	na
	Drugs	0.151	1	na	na	na	na	na
	Work style	0.1321	1	na	na	na	na	na
	Conscientious eating	0.0943	1	nc	nc	0.65	0.65	0.65
	Exercise	0.0943	0.6	0.6	0.6	0	0	0
	Cholesterol issues	0.0943	0.65	na	na	na	na	na
	Body weight	0.0943	0.66	nc	nc	nc	nc	nc
	composite utility value		0.897	0.85	0.85	0.76	0.76	0.619
	SG			90	90	85	80	65
3								
	Mental Health	0.96	1	na	na	1	1	na
	No chronic illness	0.036	0	1	1	0	0	0
	Active Life	0.0024	0	0.6	0.6	0.6	0	0
	Obsessiveness	0.00077	1	na	na	na	na	na
	composite utility value		0.961	0.998	0.998	0.962	0.961	0.961
	SG			70	50	18-20	1	1

Figure 3. Sample results from three interviews. Labels for the attribute scales arbitrarily reflect only the healthy extreme, not the unhealthy extreme. The third column contains the relative weight in a linear model. Health state zero (HS0) contains the level assigned to their current health state. Columns for the scores of five health states (HS1-HS5) contain the value assigned to five ordered levels of permanent disability following stroke. The symbol 'na' stands for "not applicable" or "no change expected" in which case the value for the current state (HS0) should be used to compute a health state's score. The composite utility value is the summary measure given by the sum of the individual attribute values multiplied by their corresponding coefficients from the third column. SG is the utility value assessed using the standard gamble method for the same health state in the same patient.

to complete the task thus far. Representative data resulting from these interviews are presented in Fig 2 & 3. Data presented in this paper are only intended to serve as examples of output from the earliest implementation of the proposed methodology. These early results illustrate in patient #1 a simple outlook on life with only three dimensions to health that also happen to be relatively equal in weight. Rating himself at the highest level in each

attribute, he thinks he is as health as is humanly possible. Patient #2 is more interesting with eight attribute scales which cluster at varying relative weights in the overall model of healthiness. It is interesting to note that patient #2 has only scored herself below the highest extreme in attributes she considers less important. It is also interesting to note the similarities and differences between the scores for post stroke disabilities using the individualized multiattribute approach and the standard gamble method. Patient #3 does not regard his health state as anything above the bottom in two of four scales. Noting labels of the poorest extreme for those scales clarifies that he is suffering chronic illness and is forced to lead the most sedentary life. This patient did reveal that he was terminally ill. Note, however, that he has a high regard for the Mental health for which he scores his state as optimal and the effect this has upon the utility scores. These scores seem high at face value but still the ranking and relative distance between scores could be interpreted as sensible. The standard gamble assessment has no way of revealing such potential detail.

Eliciting such a multiattributed utility model takes significantly more time than the direct assessment of summary measures in traditional methods. Interviews eliciting bipolar constructs alone took approximately 20 minutes per person on average. Interviews eliciting the entire value system along with a few demographic questions, four questions regarding the properties of the attributes, self assessment, and application of the resulting schema to score 5 levels of disability following stroke takes approximately 4 hours per person on average. These lengthy interviews were split into a few 30-120 minute sessions at the convenience of the volunteer. In the few interviews conducted to date, no one has been unable or unwilling to complete the interviews and all have endorsed the purpose and result of the interview when finished. However, no attempt has been made to perform the interview in a single session. Four such interviews have been completed as of this writing in which there were 3, 4, 4 and 8 attribute scales requiring 75, 110, >300 and 195 min respectively.

Discussion

It is notable that the elicitation method does not depend upon a specific health state and thus is not bound to single use. In so far as the patient can be considered (or assessed with repeated application of the instrument) to retain the same perspective, the elicited attribute scales and recomposing model coefficients are re-usable. If the patient perspective does evolve, automation makes repeat evaluation more convenient.

The traditional approaches to utility assessment have enabled codification of those systematic protocols (Lenert et al 1995) (Sumner, Nease, and Littenberg 1991). Separate programs have been written to implement the elicitation of attributes with role triads for what is called repertory grid analysis (Sewell et al 1992). The systematic nature of this novel approach would suggest similar codification for each of the component parts is feasible and their integration into a coherent application equally so. Efforts to implement this individualized utility assessment with a single computer program are underway. The discipline of codification not only will free the elicitation procedure from some constraints, it continues to more rigorously systematize the procedure. By enabling practice, experimentation and repetition, programming renders results more reliable and less vulnerable to omission.

To be complete, the model must include all factors of relevance to the interviewed subject. By employing multiple triads of acquaintances and exhausting the ideas suggested by each, the list of attributes should be complete. By allowing a response, free of acquaintance association, to terminate the solicitation of relevant factors, the risk of omitting important issues for which no example exists within the individual's familiar circle is minimized. Further, if the interview explicitly asks whether each attribute scale covers all extremes of its dimension, claims of descriptive comprehensiveness are justified regarding the individual's current perspective. It may be argued that anything missing is no more likely part of the patient's natural decision process than it is likely to be brought to mind in an interview of this extent. The elicitation is dependent upon a recognition of the patient as an authority on what is important, and what is not, which is most appropriate in regard to personal preferences.

While it is recognized that some people fail to take a course of action consistent with their expressed values, it is hoped that better understanding the subcomponents of their value system will suggest rational hypotheses for their inconsistency or failure to act. This multi-attribute approach provides an opportunity to see how individual factors may cancel each other out or create tension. It would be misguided to think that the result will not include unimportant dimensions on occasion and omit others of relevance in the mind of those outside the individual. It is the descriptive nature of the method's results that will make explicit the difference between this individual and others and, only after such explication, enable comparison and prescriptive action. Any personal inconsistencies that do exist will also be more evident and understood after this systematic elicitation.

The linear model used to compute composite scores for health states in this multiattributed representation is not the only possibility. Varied appealing alternative models which are more complex are discussed in the literature of utility assessment and decision theory (Wellman 1985). The approach put forth here is simply intended as a baseline from which more complex models might be pursued with presumably more complex elicitation protocols. It is

reasonable, however, to assume that such embellishments would add to the time cost for elicitation and their benefits should be weighed against that marginal cost. It follows that lessons learned in the codification of the linear model based approach would apply to that of more complex models.

Validation

The issues which must be addressed to validate this method of utility assessment include the usual psychometric cadre of feasibility, reliability, comparability, construct validity and criterion validity. Feasibility as expressed by the ability of individuals to complete the assessment protocol and generate output is promising as indicated previously. Reliability and comparability evaluations as expressed by measuring the correlation of results between repeated assessments as well as between this and traditional methods are planned. The sample data illustrates interesting possibilities. Construct and criterion validity as expressed by analysis of the successful prediction of related behavior and relation to other health status measures could be evaluated with longer ranged studies.

Opportunities to embellish with AI

Automating the elicitation procedure of individualized utility assessment would greatly ameliorate the increased time cost of this proposed methodology. Validation studies would be made much easier for the sake of reduced variables if programming the interview is successful. Although the proposed methodology is systematic, the queries involved are dynamically generated based upon the specific terms elicited during the interview. This renders a pre-printed survey or scripted protocol inadequate for this purpose. There are several aspects to this methodology which suggest that various tools familiar to the artificial intelligence community would be beneficial to apply. For instance, although the interview is very structured it is not rigidly structured. The structure may be loose enough to warrant a rule based control paradigm, e.g., a number of conditions can indicate when one triad of people are no longer producing fresh ideas, or when to switch between titrating and bracketing for progressively changing the probabilities in standard gamble steps. Justification for employing additional artificially intelligent embellishments arise when the protocol embraces past responses to facilitate current elicitation. To do so requires self-organizing storage and retrieval so as to enable real-time aide without slowing the interview process. Semantic networks are suggested by the idea that attribute scales might be compared to, checked by or built on the seeds of previous responses. Well thought out algorithms could minimize bias and yet assist the person struggling to articulate an idea they definitely want to assert. Similarity measures are required to compare and rank the proximity of interview responses to previous responses. It follows that while natural language processing does not provide a basis for similarity measures in the time frame of these interviews, the principles of natural language processing coupled with constraint propagation could contribute to such measures or serve to maintain working semantic networks off line.

As with any interactive software, user interface issues play a role. In a domain where research indicates that the values elicited by a vertical rating scale will systematically differ from the same instrument with a horizontal rating scale, the introduction of more steps with graphic displays, mouse events and graphical editing will raise more questions. Little has been done to empirically address these complex issues while technical capacities continue to soar. It is unclear whether a novel approach to elicitation should wait for empirical evidence to make user interface decisions. As elicitation instruments grow in complexity to ameliorate certain issues, the number of steps involved in the paradigm of typed-in text to textual prompts can become flogging. Intuitive interactions using mouse and graphics may be required to avoid overload. In this case it may be easier to ask the user to drag an intermediate attribute scale label from a starting location and drop it on a line relative to the two extremes of that attribute scale than it is to properly word explicit instructions for giving a number between 0 and 1 for that same purpose. The process may be improved further if it is conducted by a program which tracks the time taken to respond to different forms of the question and responds accordingly. These are all promising directions for research following the establishment of a complete first prototype.

Conclusion

A method for eliciting relevant dimensions of health from an individual patient in their own terms has been demonstrated. The approach is systematic. The list of attributes is made complete by exhausting the ideas brought to mind by real relationships. By insuring the extremes are covered for each scale in that complete list, the multiattribute schema is as comprehensive as the patient's grasp of the perceived world. Where that is deficient in anyone's view becomes explicitly clear. This result pointedly prescribes specific demands for communication and subsequent understanding. The results are operational for both the patient who may not otherwise think through the issues and for the provider who is empowered to think and express instructions, recommendations and consequences in the patients terms and world view. The scales and coefficients output by this method of utility assessment are far less dedicated to a solitary context than the customary summary measures. Their re-use can serve to distribute the cost of elicitation, be that in hours, dollars or

some other economy. This systematic approach also suggests a promising potential for automation owing to the fact that the components of the protocol have been programmed. Automation further spares the cost of and constraints upon decision analysis. Success at codification may depend heavily upon implementation of recognizable elements from artificial intelligence. However, to the degree with which the protocol can be automated it will enable individuals to reveal their values in the form of utility assessments outside the limitations of highly trained analysts in the comfort and privacy of their home or anywhere a computer terminal may be used. It follows that this freedom to express the preferences peculiar to an individual would allow more frequent assessment, even experimentation. Like the now famous spreadsheet model, this would allow the patient to explore the meaning and ramifications of their assertions and variations thereof, making more certain their approval of the final representation of their desired basis for judging health states. Patients could "take home" the values clarification tool and "toy" with it until they were comfortable with the result. It then follows that providers would have more confidence in the values expressed. This, coupled with favorable psychometric analyses would result in more confident application of decision analysis on an individual basis and a consequent increase in satisfaction with decisions. If nothing else, this method does for utility values what decision analysis does for treatment options, it makes the issues involved explicit on an individual basis fostering better communication in more familiar terms.

Acknowledgments:

The author is supported by a Medical Informatics Training Fellowship from the National Library of Medicine under grant No. 2T15LM07092.

References

von Neumann J.; and Morgenstern O. 1976. *Theory of Games and Economic Behavior*. Princeton: Princeton University Press.

Eckman, M. 1997. Personal Communication. Division of Clinical Decision Making; New England Medical Center, Boston.

Fransella, F. and Bannister, D. 1977. *A Manual for Repertory Grid Technique.* London: Academic Press.

Ryle, A. 1975. *Frames and cages: the repertory grid approach to human understanding.* London: Published for Sussex University Press by Chatto and Windus.

Keeney R.L. and Raiffa H. 1976. *Decisions with Multiple Objectives: Preferences and Value Tradeoffs*. New York, NY: Wiley & Sons.

Lenert L.A.; Michelson D.; Flowers C.; and Bergen M.R. 1995. IMPACT: An Object-Oriented Graphical Environment for Construction of Multimedia Patient Interviewing Software. In Proceedings of the 19th Annual Symposium on Computer Applications in Medical Care, 319-323. New Orleans, LA.: American Medical Informatics Assoc.

Sumner W.; Nease R.; Littenberg B. 1991. U-titer: A utility assessment tool. In Proceedings of the Symposium on Computer Application in Medical Care, 701-705. Washington, D.C.: American Medical Informatics Assoc.

Wellman M.P. 1985. Reasoning About Preference Models. MS Thesis, Dept. of Electrical Engineering and Computer Science, Massachusetts Institute of Technology.

Sewell K.W.; Adams-Webber J.; Mitterer J.O.; and Cromwell R.L. 1992. Computerized repertory grids. *International Journal of Personal Construct Psychology* 5:1-23.