

# Acceptability of Computerized Visual Analog Scale, Time Trade-off and Standard Gamble Rating Methods in Patients and the Public

Leslie A. Lenert M.D., M.S., and Ann E. Sturley, DrPH, BSN  
Veterans Administration San Diego Healthcare System, San Diego, California

## ABSTRACT

*One technique to enhance patient participation in clinical decision making is formal measurement of preferences and values. Three commonly applied methods are a visual analog scale (VAS), the standard gamble (SG), and the time trade-off (TTO). We studied participants subjective experience using computer implementations these methods using scale we call the VIBE (for Value Instrument Battery--Evaluation) that measures four aspects of user acceptance (clarity, difficulty, reasonableness, and comfort level) Studies were performed in two groups: patients with HIV infection (n=75) and a convenience sample of the general public (n=640). In the patient study, VIBE scores appeared reliable (Cronbach's alpha of 0.739, 0.826, and 0.716, for VAS, SG, and TTO ratings, respectively.) Patients' acceptance of the VAS the highest, followed by the TTO and the SG method ( $p < 0.05$  for all comparisons). Despite significant enhancements in computer software for measuring SG preferences, observed differences in acceptance between SG and VAS methods were replicated in the general public study ( $p < 0.0001$  for differences). The results suggest developers of clinical decision support systems should use VAS and TTO rating methods where these methods are theoretically appropriate.*

## BACKGROUND AND INTRODUCTION

Decision analysis is a method for rationale decision making which uses mathematical formulae to weigh the probabilities of different outcomes and thus the risks and benefits of decisions. Beginning in the early 1970's researchers first began to apply decision analysis to the health field. A long-standing goal of researchers has been to develop practical ways to apply this powerful philosophy in patient care. One obstacle has been the difficulty of measurement of values or utilities for health outcomes. These subjective judgments of the desirableness of different outcomes are necessary to assess the risk-benefit ratio of different treatments. Recently published work examining the use of preference elicitation methods in decision

support systems includes work by Prothro et al.(1) and by Ruland (2).

There are several different ways to measure how highly individuals value health outcomes. The approach that is conceptually most consistent with the theories underlying decision analysis is the standard gamble (SG) (3). In the SG, an individual determines the point when he or she is indifferent (e.g., cannot chose between) a certain outcome and a gamble with some probability of a better outcome (in health applications, most often "perfect" health) and 1 minus that probability of a worse outcome (in health applications, often "death"). In health contexts, the indifference point can be approximated as the maximum risk of death that a person is willing to accept for a treatment that cures the condition being studied.

Interviews to measure SG ratings are either performed using automated computer interviewing methods or by specially trained research assistants. In addition to being conceptually difficult, the process is labor intensive if administered by an interviewer. Recent work has shown that it is feasible to use computer methods to obtain standard gamble utility ratings over the Internet (4) and obtain predictions from decision models over the Internet (5). Thus, it might be feasible to combine these methods in a decision support system. Many of the technical considerations for linking the two technologies have been overcome (6).

Other approaches that measure values include the time trade-off (TTO) and visual analog scales (VAS). Both methods produce results that are numerically different from the gamble(3). The differences can be large and might result in an different (potentially incorrect) answer if used in a decision analysis. In health applications, VAS value measurement is performed by having the subject determine the relative distance of the state in question from two anchors conditions, typically death and perfect health. In the TTO, the rater determines whether living a shorter period of life in good health is better than living a longer period in ill health and, if so, the length of life in good health

that is equivalent to a longer period with the health condition under study. The TTO rating is the ratio of the former to the later. The TTO method was developed specifically to simplify the task of preference measurement, however, it has not been proven to be better accepted by patients.

As is obvious from the above descriptions, the gamble method and the time trade-off much more complex than the analog scale method. While some have suggested approaches for converting analog scale ratings into approximations of the gamble there is presently no viable way to do this (7) accurately at an individual level. Thus, would-be developers of decision support systems are faced with the choice of using the more complex standard gamble and time trade-off or finding creative ways to use the simpler analog scale. The objective of this paper is to study users' subjective evaluation of each of these measures, to learn how best to tailor such systems to users' needs and preferences.

Despite growing use of computer and interviewer-based methods for decision support, relatively little work has been done to better understand users' responses to different value measurement methods. Most previous studies have limited subjective evaluation of users' responses to the impression of the interviewers' supervising data collection. However, the scope of this issue is broader. Bergen et al. proposed four different aspects for subjective evaluation of a preference elicitation method (8). These are:

- The *clarity* of the rating task
- The *difficulty* of the rating task
- The *reasonableness* of the rating task
- The *comfort level* of the individual in using the task for medical decision making.

Clarity, difficulty and reasonableness are conceptually distinct concepts. Tasks can presumably be clear but difficult, or clear but unreasonable. Ultimately, however, the central issue is how comfortable a person would be using a method in medical decision making. Contemporary theories of validity focus on the validity of measures for their intended use. An individual who is not comfortable using a measure for individual decision making may not believe that the measures represent his or her preferences, and hence the validity of the measures is called into question. To explore these issues, we have developed a user acceptance questionnaire for preference elicitation instruments we call the VIBE

(for Value Instrument Battery—Evaluation). In the VIBE, each factor in the Bergen et al. model is evaluated using a four point Likert type scale, that required subjects to chose between a positive and a negative evaluation. For example, when rating the comfort using a method in decision making, participants could respond “Very comfortable” (4 points), “Comfortable”, “Uncomfortable” or “Very uncomfortable” (1 point). In addition, subjects' understanding of the conceptual model underlying each method was explored with open-ended questions.

As part of efforts to validate a iPACT3 program (9), a web based tool for design, construction and administration of health value measurement protocols, we administered the VIBE to participants in two studies. In the first experiment, we studied the values of HIV infected patients and for hypothetical health states with adverse effect of HIV medication called the lipodystrophy syndrome. In the second experiment, we studied the values of general public for states with schizophrenia. In both cases, participants viewed and used rating methods created with the iPACT3 web site.

## METHODS

The HIV study was conducted in a urban university hospital HIV clinic in patients infected with the Human Immunodeficiency Virus (HIV). A research assistant met with trained them in use of the computer, and then started them on completion of the survey. He returned to assist users at their request or if judgment suggested they were having difficulty operation of the computer or rating tasks. After training individuals in use of analog scale methods and standard gamble methods, subjects rated three health states—a largely asymptomatic person with HIV infection, a person with HIV infection complicated by an adverse drug effect, but otherwise asymptomatic, and their current health. Descriptions of the health states were in text format, with photographs that showed the effects of the lipodystrophy syndrome. They then rated each state, first using the VAS, then the SG, and last the TTO.

At the completion of the computer survey, patients completed the VIBE questionnaire on paper. To evaluate the internal consistency of the acceptability questionnaire, we calculated Cronbach's alpha for SG, TTO, and VAS instruments.

Based on the feedback from the HIV study, we adapted the computer instruments to enhance their appearance and simplicity. Focusing on standard gamble and analog scale instruments, we simplified the screen design, the SG assessment procedure, and enhanced training materials with explanations on how to use the instruments in an effort to improve acceptance of the gamble. Training materials were converted to Flash 4 (Macromedia Software) animations, that demonstrated, in a step-by-step fashion, how to operate the instruments. In addition, we used a simpler procedure to measure standard gamble utility, where subjects slowly increased the risk in standard gamble ratings to their indifference point

The second survey explored preferences for health outcomes in schizophrenia. Participants in the survey viewed six patterns of symptoms of schizophrenia portrayed using digital video materials. They then rated each state with the VAS and SG rating methods. At the end of the survey, they completed the VIBE questionnaire. Participants evaluated each aspect of the VAS and SG rating methods using the same four-point Likert type scales used in the previous experiment, except they completed the forms "on-line".

## RESULTS

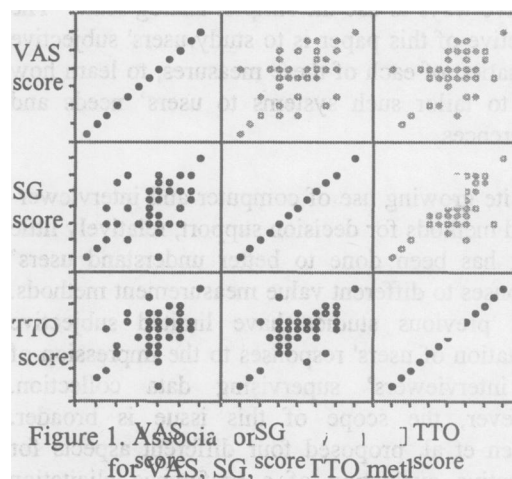
A total of 75 HIV patients completed the HIV survey. Patients were relatively well educated (80% with some college education), ethnically diverse (35% minorities), and had a median age of 40 to 50 years. Eighty-nine percent were receiving anti-retroviral therapy. All 75 enrollees completed the computer.

The ratings performed using VIBE appeared reliable, with Cronbach's alpha of 0.739, 0.826, and 0.716 for the VAS, SG, and TTO respectively. The VAS was rated more favorably than the SG (95% CI for difference between methods 1.7 to 0.59) and somewhat more favorable than the TTO (95% CI for the difference 0.046 to 0.107.) The TTO was rated more favorably the SG (95% CI for difference of 0.12 to 0.162.) There was substantial variation within individuals in ratings which method was the most favored, as shown in Figure 1.

Looking specifically at responses to the item on comfort with use in decision making, there were strong trends toward a greater comfort with VAS and TTO methods over the standard gamble (p=0.056 and 0.089, Wilcoxin sign rank test).

Patients appeared equally comfortable with VAS and TTO methods (p=0.70). Results are summarized on the left side of Figure 2.

In the second study, delivery of 30,103 email invitations to survey panel members resulted in 1302 study enrollments and 640 completed questionnaires, with much of the drop attributable to lack of access to streaming video software required for viewing of the health states in the survey. Subjects completing the survey were well educated (80% with some college education) with a median age of 35 to 44 years. Subjects had a diverse ethnic background (42% minorities (achieved through over sampling of minority groups in the panel.)



The computerized VIBE was slightly less reliable with alpha values of 0.64 and 0.77 for the VAS and SG respectively. Results, otherwise, were similar to the first study. The average rating for the VAS scale was 1.5 units higher (95% CI 1.7 to 1.4, p<0.001, paired t-test.) There was considerably heterogeneity among individuals in acceptance, with some individuals rating the gamble more favorably than the analog scale (r = 0.501 Spearman correlation coefficient). Looking at the critical question of comfort in use in decision making, participants were more comfortable using the VAS than the SG in decision making (mean difference 0.43, p <0.001 Wilcoxin sign rank test.) The distribution of comfort level is shown in the right half of Figure 2.

The larger sample size of the second study supported testing for associations between the

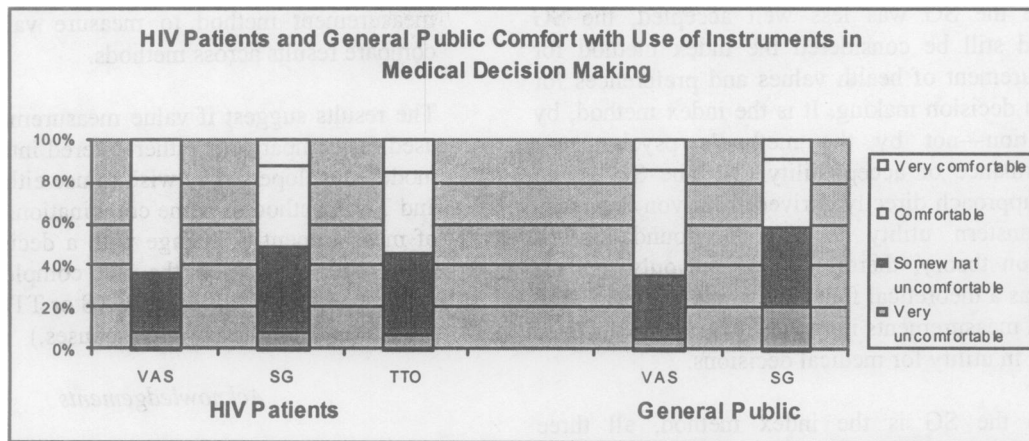


Figure 2. Comfort with using VAS, SG, and TTO methods for clinical decision making.

VIBE score and demographic features of the population. No significance difference in acceptance scores was seen across ethnic groups or educational level in acceptance of the VAS or the SG rating methods. Interestingly, men appeared to rate both methods more highly, with statistically significant differences VIBE scores for the SG and a trend in differences in ratings of the VAS method as shown in Figure 3.

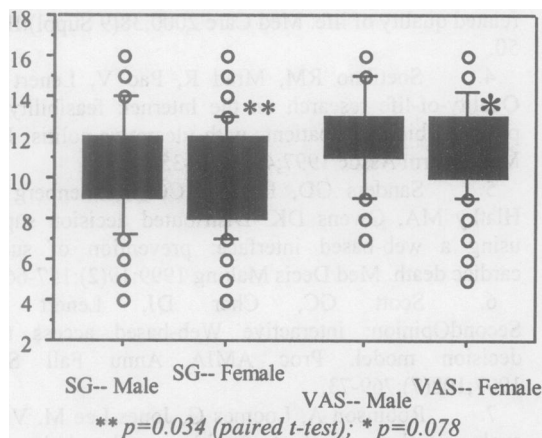


Figure 3. Differences in acceptance of SG and VAS methods across gender.

## DISCUSSION

In this paper we report on the use of a four-item questionnaire, the VIBE, to evaluate user acceptance of three different computer methods for measurement of health values. Comparisons were performed in two relevant populations—patients with HIV disease and in convenience sample of

members of the Internet using public. Results of both studies were consistent. Study participants rated the analog scale rating methods most highly. There were small differences in over all the acceptance score between the analog scale and the time trade-off and larger ones between the analog scale and the standard gamble. In both patients and the public, the respondents were most likely to be comfortable using the analog scale for decision making. Statistically significant results were seen in the larger general public sample and strong trends were seen in the patient study.

To measure participants' acceptance of different rating methods we used the VIBE. This brief instrument looks at four aspects of the experience of using a preference measure. In addition to the questions, the VIBE has a picture of each rating instrument to remind subjects of the measure during their rating tasks. Results revealed that the VIBE has acceptable reliability and appears to discriminate between rating methods. The VIBE may be best administered in paper form, after completion of a survey, as reliability estimates were slightly higher in this format.

The simple four-question VIBE scale appears to be a useful tool to evaluate patients and the public's meta-preferences procedures for measurement of their values. Through use of such tools protocol and software developers may be able to refine procedures for measurement of the SG and other procedures so that acceptance nears that of VAS or TTO methods. Researchers interested in using the VIBE scale can download a sample version at <http://preferences.ucsd.edu/vibe>.

While the SG was less well accepted, the SG should still be considered the index method for measurement of health values and preferences for use in decision making. It is the index method, by definition—not by the method's psychometric performance or acceptability (3). The SG is the only approach directly derived from von Neuman-Morganstern utility theory—the foundation of decision theory; therefore, it is the only method that has a theoretical foundation that can justify the use of measurements in models to estimate gains or losses in utility for medical decisions.

While the SG is the index method, all three methods measure health values (how good or bad someone feels a state is) (3). Both the SG and the TTO *preferences*—and can be used to determine which state or condition an individual *prefers* (as some trade-offs is required to determine if someone *prefers* one thing to another.) In addition, the SG captures an individual's risk attitude, or willingness to assume risk to gain benefit. A large body of literature has shown that measurements performed using one assessment method, differ from those performed with some other method (3).

Because all three methods measure health values, all produce results that are potentially clinically relevant. In some circumstances, the TTO may be the most relevant measure. If the objective is to determine whether to forgo some period of life expectancy brought about by therapy with quality of life effects (such as chemotherapy for cancer), the TTO is particularly salient. Other medical questions require understanding only of which state patients see as less severe. For example, recent work by Ruland described use of patients' VAS ratings to help prioritize nursing care interventions (2). Thus, practical clinical tools might be built using any of the rating methods.

While measurement of values using one rating method may be all that is needed to set priorities, or to get input for a decision model, if the decision is an important one, value measurements may need to be performed using two or more methods. Failure to maintain a consistent rank order across different methods of assessment of values (also called failure to satisfy procedural invariance of preferences) is a not uncommon finding in preference measurement studies (3). It appears to be due to failure of at least one of the preference measurement methods applied to successfully measure values (3). Therefore, it is important for clinical systems to employ more than one

measurement method to measure values and to compare results across methods.

The results suggest if value measurements will be used for comparisons rather entered into a decision model, developers may wish to use either the VAS and TTO method or some combination. If the goal of measurement is linkage with a decision model, may need to include the SG complemented by simpler and better-accepted VAS or TTO (to assess the internal consistency of responses.)

#### *Acknowledgements*

This work was funded by grants from Janssen Pharmaceuticals and GlaxoSmithKline Pharmaceuticals.

#### REFERENCES

1. Protheroe J, Fahey T, Montgomery AA, Peters TJ. The impact of patients' preferences on the treatment of atrial fibrillation: observational study of patient based decision analysis. *Bmj* 2000;320(7246):1380-4.
2. Ruland CM. Decision support for patient preference-based care planning: effects on nursing care and patient outcomes. *J Am Med Inform Assoc* 1999;6(4):304-12.
3. Lenert L, Kaplan RM. Validity and interpretation of preference-based measures of health-related quality of life. *Med Care* 2000;38(9 Suppl):II138-50.
4. Soetikno RM, Mrad R, Pao V, Lenert LA. Quality-of-life research on the Internet: feasibility and potential biases in patients with ulcerative colitis. *J Am Med Inform Assoc* 1997;4(6):426-35.
5. Sanders GD, Hagerly CG, Sonnenberg FA, Hlatky MA, Owens DK. Distributed decision support using a web-based interface: prevention of sudden cardiac death. *Med Decis Making* 1999;19(2):157-66.
6. Scott GC, Cher DJ, Lenert LA. SecondOpinion: interactive Web-based access to a decision model. *Proc AMIA Annu Fall Symp* 1997;108(7):769-73.
7. Robinson A, Loomes G, Jones-Lee M. Visual analog scales, standard gambles, and relative risk aversion. *Med Decis Making* 2001;21(1):17-27.
8. Bergen M, Parineh M, Garber A, Flowers C, Lenert L. Acceptability of multimedia standard gamble and willingness-to-pay preference elicitation procedures. *Proceedings of Society for Medical Decision Making* 1995;17.
9. Lenert LA. iIMPACT3: online tools for development of web sites for the study of Patients' preferences and utilities. *Proc AMIA Symp* 2000;37(2).