Volume 6 • Number 4 • 2003
VALUE IN HEALTH

Assessment of Respondent Acceptability of Preference Measures: Discriminatory Power of Graphic Positioning Scale versus Traditional Scaling Measures

Duska M. Franic, PharmD, PhD,[1] Dev S. Pathak, DBA[2]

[1]College of Pharmacy, The University of Georgia, Athens, GA, USA; [2]College of Pharmacy, The Ohio State University, Columbus, OH, USA

ABSTRACT

Objective: The objective of this study was to compare discriminatory power of two different measures—graphic positioning scale (GPS) versus traditional scale (TS)—in assessing respondent acceptability of three preference measures: visual analog scale (VAS), standard gamble (SG), and willingness to pay (WTP).

Methods: Two face-to-face interviews were conducted at least 1 week apart in a convenience sample of women aged 22 to 50 years with no history of breast cancer or cancer requiring chemotherapy. Study participation required completion of two surveys: one evaluating health preferences for an acute condition (chemotherapy-induced nausea and vomiting) and one evaluating a chronic condition (breast cancer). Data were collected from March 2000 to June 2000 at Ohio State University. Respondents were randomized to either GPS or TS surveys. Data analysis was a two-step process. First, a four-way multivariate repeated-measures analysis of variance (MANOVA) was conducted to assess respondent acceptability of three-preference metrics—VAS, SG, and WTP—in health-care decision making. Each of the four dependent variables, difficulty, clarity, reasonableness, and comfort in use in decision making, was measured on 9-point Likert scale. Second, a mixed design univariate analysis of variance (ANOVA) was performed for each dependent variable to optimize MANOVA analysis. Univariate ANOVAs, $2 \times (2 \times 3)$, were composed of three independent variables: assessment (GPS/TS), condition (acute and chronic), and preferences (VAS, SG, WTP).

Results: Of 126 respondents, 119 were usable and complete. MANOVA results showed ($P < .05$) for two main effects, condition ($F_{4,114} = 6.375$) and preferences ($F_{8,110} = 9.290$), and two significant interactions, condition × assessment ($F_{4,114} = 3.421$) and condition × preferences ($F_{8,110} = 2.087$).

Conclusion: GPS has higher discriminatory power than TS in assessing respondent attitudes toward health preference measures. Results showed that respondents had more difficulty and less comfort when making decisions for chronic than for acute conditions. Results also show that respondents regard WTP as a more reasonable decision-making tool when assessing acute interventions in preference to SG and vice versa for chronic conditions. Of VAS, SG, and WTP methods, VAS was perceived as being the easiest to understand. These results can be explained by direct versus indirect comparisons made with GPS and TS methods, respectively.

Introduction

A desirable property of survey design is brevity. Surveys should be of sufficient length to include all important items but not so time-consuming to result item omission, nonresponse, fatigue, or satisficing [1–3]. Satisficing can occur when an individual is faced with multiple options. Careful review of all options can be quite labor-intensive and it has been suggested that people take shortcuts by using "rules of thumb." That is, instead of carefully weighing all the choices and selecting the optimal alternative, the individual selects the first option that appears reasonable based on some minimal criteria for acceptability. Therefore, satisficing means the individual did not necessarily select the best option. This is believed to be a potentially serious problem in long surveys [3].

Shorter questionnaires are more likely to result in higher response rates. Several suggestions have been made to reduce survey length such as only including pertinent items. However, resultant number of

Address correspondence to: Duska M. Franic, Assistant Professor, College of Pharmacy, The University of Georgia, RC Wilson Building Room 254, Athens GA 30602-2354. E-mail: dfranic@rx.uga.edu

 1098-3015/03/$15.00/483 483–493

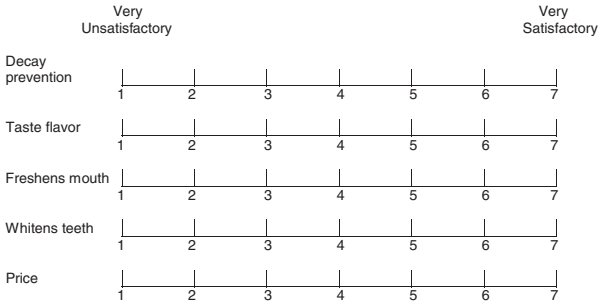


Figure 1 TS measuring toothpaste satisfaction. One set of items required for each brand tested in study. (Adapted from Narayana [4].)

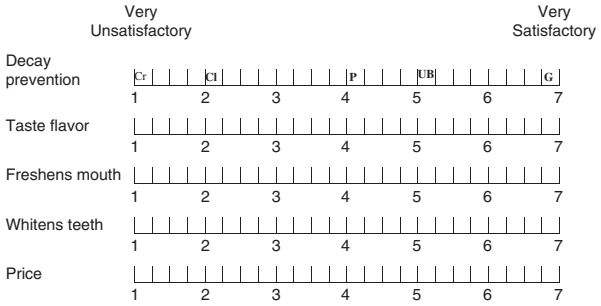


Figure 3 GPS measuring toothpaste satisfaction. (Adapted from Narayana [4].)

reduced items may result in loss of data. Narayana [4] reported in 1977 successfully reducing survey length without loss of information in an attitudinal market survey using graphic positioning scale (GPS) as an alternative to the traditional scale (TS).

TS surveys require presenting one item per question; therefore, a 25-item survey would require 25 questions. For example, a marketing study designed to measure the respondent satisfaction for different toothpaste brands randomly assigned half the respondents to TS and the remainder to GPS [4]. Information was elicited regarding five different attributes for all brands. TS required respondents to answer the same set of 5 questions for each brand totaling 25 items for satisfaction dimension (Fig. 1). In contrast, the GPS can provide information on all 5 brands with only 5 instead of 25 items (Fig. 2). For example, for the five brands tested in the study, Crest® (Cr), Colgate® (Cl), Pepsodent® (P), Gleem® (G), and Ultrabrite® (UB), respondents would rate each satisfaction question by marking the scale with the appropriate brand abbreviation (Fig. 3). Therefore, GPS requires each respondent to directly compare each of the five brands without repeating each question five times and thereby reducing survey length [4].

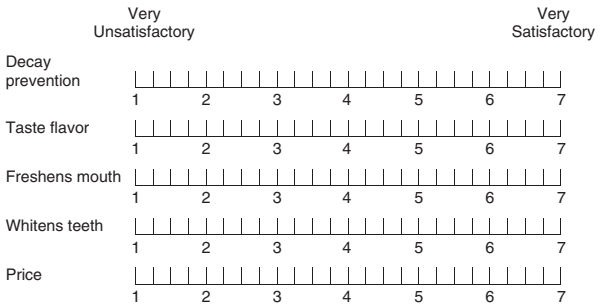


Figure 2 GPS measuring toothpaste satisfaction. (Adapted from Narayana [4].)

In the marketing study, Narayana [4] also collected data on variable importance and overall ratings for each of the five brands of toothpaste. The author regressed overall ratings for TS and GPS methods onto perceived satisfaction and importance for each of the five brands. Analyzes showed GPS to be a better predictor than TS accounting for a greater proportion of the variance. The author concluded GPS required respondents to make direct comparisons of each brand for each attribute (Fig. 3) of interest while comparisons made with TS were indirect (Fig. 1). In contrast, the TS approach placed comparison focus on the brand not the attribute—the respondent was required to complete the survey for one brand before considering the same attributes on the next. Narayana [4] suggested that real-life decision making is typically based on individual attribute comparisons, not brands, and concluded that the use of GPS over TS generated more discriminatory power implying better data quality [4].

Narayana also acknowledged that the differences observed could have been due to the "graphic" scaling used in GPS versus TS (note the additional rule marks in Fig. 2 over those of Fig. 1). The purpose of this study is to build on the work of Narayana by using the same graphic scaling system for both GPS and TS, so any differences cannot be accounted for by disparity in the appearance of the "ruler." Additionally, this study uses a health-care setting as opposed to a marketing setting. The primary focus of this study is to evaluate the potential for reduction of questionnaire items without loss of data in a health outcomes survey using GPS versus TS for both acute (temporary) and chronic ("rest of life") conditions. Definitions for acute and chronic conditions used in this study were based on Torrance et al. [5–10] and the US National Center for Health Statistics [11]. For the purposes of this study, an acute condition is defined as a severe, intense con-

dition of short duration—less than 3 months. For example, chemotherapy-induced nausea and vomiting (CINV) is brief, lasting an average of 3 days, and is intense. Chronic health states describe long-term conditions, lasting the rest of a person's life. In other words, they are permanent until death conditions [5–10]. The US National Center for Health Statistics in 1994 referred to chronic conditions as those lasting a long time—greater than 3 months and typically implying a low intensity [11]. For the purposes of this study, a chronic condition is defined as a health effect of long duration (i.e., greater than 3 months). For example, in patients with breast cancer, although often fatal and not necessarily curable, the health state usually lasts longer than 3 months.

Currently, decision analysis literature recommends preference assessment using visual analog scale (VAS), standard gamble (SG), and willingness to pay (WTP) methods for both temporary and permanent conditions; however, the applicability of these three preference measures for acute and chronic conditions has been questioned [12–15]. The secondary focus of this study is to determine the applicability of these three metrics in acute and chronic conditions.

Objective

The purpose of this article is to compare the discriminatory power of two different measures—GPS and TS—in assessing respondent attitudes toward three preference measures, VAS, SG, and WTP, in a health outcomes survey.

Methods

Participants

Target population was women aged 22 to 50 years with affiliations with a Midwestern university. Respondents were recruited using E-mails, word of mouth, and fliers. Because the larger study involved hypothetical scenarios regarding breast cancer, only women were recruited for study participation. Furthermore, to avoid possible emotional distress or discomfort, respondents with a history of breast cancer or cancer requiring chemotherapy were excluded from participation. The study presented in this article was part of a larger project reported in detail elsewhere [13–15].

Potential study respondents were informed that all information provided would remain strictly confidential. Before the study, participation respondents completed a consent form stating they agreed to participate in the study voluntarily. Respondents were free to discontinue participation at any time. The study was reviewed and approved by Internal Review Board at Ohio State University.

Each respondent in this study participated in two face-to-face interviews requiring an average of 1 hour each to complete. One interview focused on utility elicitation for an acute condition, CINV, and the other, a chronic condition, breast cancer. The purpose of the larger study was to compare different utility elicitation methods (SG, VAS, WTP) for acute and chronic conditions based on their common usage or underlying theoretical foundations in welfare economics. As part of a large study, WTP was regressed onto quality-adjusted life-years measured using VAS or SG methods [13].

To account for ordering and testing effects in the larger project, half of the respondents were randomly assigned CINV surveys first, with the remainder assigned breast cancer surveys first. Respondents were then randomized to receive GPS or TS for both interviews to assess respondent acceptability of three preference measures—VAS, SG, and WTP—used in this study. All interviews were fully structured and interviewer administered to minimize potential biases [13].

Each interview consisted of eight sections. The first section provided background material regarding respective oncology health state: CINV or breast cancer. The second section asked respondents to read and rank six health states. Respondent preferences were rated first using VAS and then SG. VAS is also known as a rating scale or feeling thermometer and is the most commonly used technique for measuring preferences because of its simplicity [16]. Many versions of the rating scale are available. The one used in this study is presented in Fig. 4, with the vertical number line with the least and most desirable health states labeled as 0 and 100, respectively [7]. SG is regarded as the "gold standard" in preference measurement because it measures preferences under conditions of uncertainty [7,16,17]. SG technique is a lottery comparing two choices. One alternative, option A, is a gamble between the most and least desirable health states with a probability of P and (1 – P), typically defined as perfect health for time horizon *T*, and immediate death, respectively. The other alternative, option B, does not involve risk, and is an intermediate health state also for time horizon *T*. The value for P is varied until the individual is indifferent between the two options in the gamble. The indifference P value, P*, is the assigned preference or utility score for the intermediate health state [7,14]. The third section recorded respondents' sociodemographic information. The

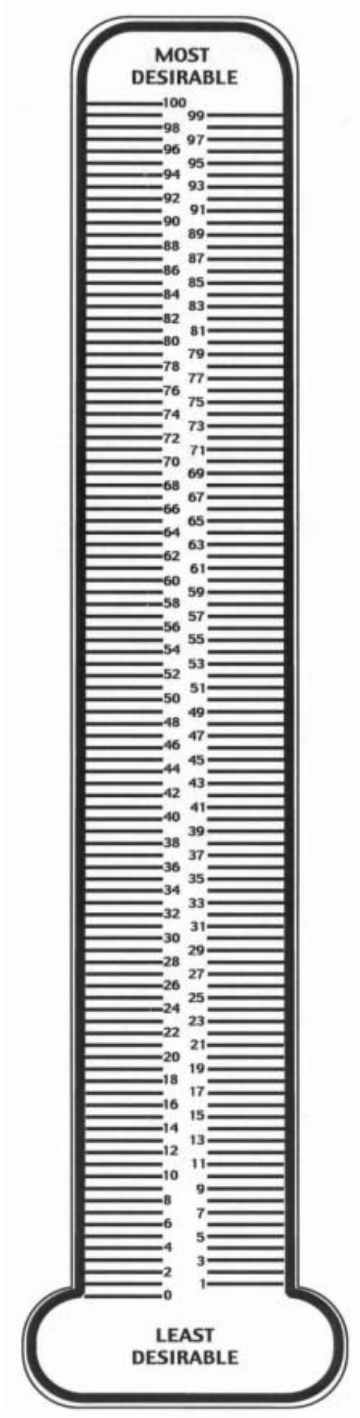


Figure 4 VAS.

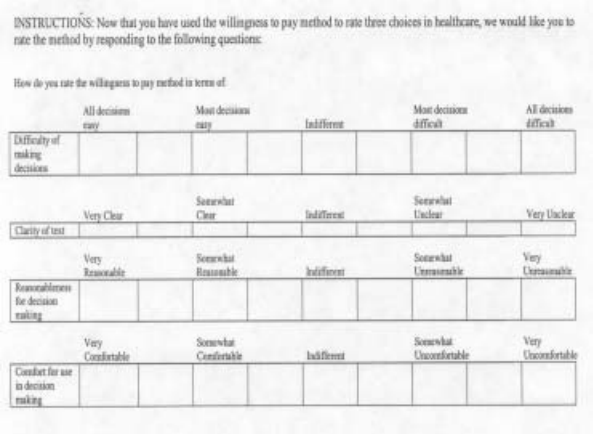
fourth section assessed WTP for health-state scenarios used in the VAS and SG sections. WTP measures respondents' preferences or utility for health states in monetary terms, the maximum amount of money they are willing to pay to receive a health benefit [17]. The fifth section provided respondents with an opportunity to assess acceptability of the three different preference measures using GPS or TS method. The sixth section asked respondents to rate health-state scenarios used in the interview, which was part of content validity of instrument [13]. The final section was completed by the interviewer, who assessed the status of the interview. This article concentrates on the fifth section of the interview.

Design

A cross-sectional descriptive study design was used [18]. A multivariate repeated measures design—multivariate analysis of variance (MANOVA)—was conducted because four dependent variables were evaluated in the study. The dependent variables were used to evaluate respondent attitudes toward the three preference metrics (VAS, SG, and WTP) for acute and chronic conditions for TS and GPS on 9-point Likert scales. The four dependent variables were: 1) clarity of text (1 = very clear to 9 = very unclear); 2) difficulty making decisions (1 = all decisions easy to 9 = all decisions difficult); 3) reasonableness for decision making (1 = very reasonable to 9 = very unreasonable); and 4) comfort for use in decision making (1 = most comfortable to 9 = very uncomfortable) [19–21]. The midpoint of each of the four scales was labeled "indifferent."

Three independent factors included in the model were labeled assessment, condition, and preference. The first factor, assessment (the between factor), described whether respondents were randomly assigned to receive GPS or TS surveys for both interviews. The two remaining independent variables were within factors—all respondents were exposed to all levels of these two factors. The second factor, condition, was assessed on two levels and referred to the type of health state being assessed: an acute condition (CINV) or a chronic condition (breast cancer). The third factor, preference, referred to the three preference measures, VAS, SG, and WTP. The next section describes the three factors in more detail.

Assessment. Study participants randomly assigned to TS responded to four items on a 9-point scale for each of the three preference measures, VAS, SG, and WTP, for a total of 12 items, for both acute and chronic surveys, for a total of 24 items. All respondents completed the TS section of the survey in the same chronological order as preference measurement was presented, that is, VAS, SG, and finally WTP. Figure 5 presents TS items for WTP section. Respondents assigned to TS would be also required to complete TS items for VAS and SG.


INSTRUCTIONS: Now that you have used the willingness to pay method to rate three choices in healthcare, we would like you to rate the method by responding to the following questions:

How do you rate the willingness to pay method in terms of:

	All decisions easy		Most decisions easy		Indifferent		Most decisions difficult		All decisions difficult
Difficulty of making decisions									

	Very Clear		Somewhat Clear		Indifferent		Somewhat Unclear		Very Unclear
Clarity of test									

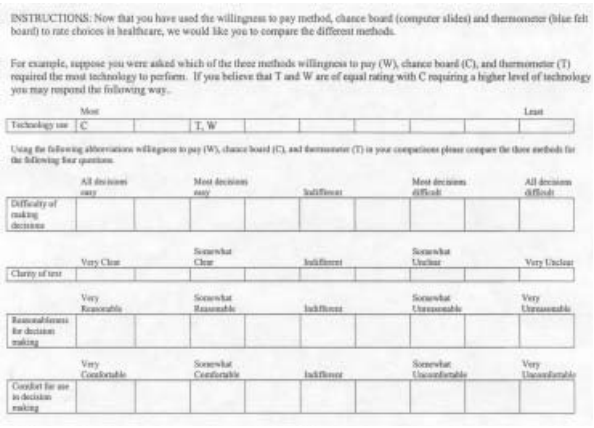
	Very Reasonable		Somewhat Reasonable		Indifferent		Somewhat Unreasonable		Very Unreasonable
Reasonableness for decision making									

	Very Comfortable		Somewhat Comfortable		Indifferent		Somewhat Uncomfortable		Very Uncomfortable
Comfort for use in decision making									

Figure 5 TS—WTP section. One set of items required for each preference method tested in study.

GPS required respondents to answer four items, but each of the four items consisted of three levels. An example of how to use GPS was provided to respondents in the beginning of GPS survey section. Figure 6 shows GPS survey used in this study (four items for both acute and chronic surveys). Figure 6 also includes the example used to show respondents how to complete GPS items.

Conditions. In this study, two conditions were assessed: an acute condition, CINV, and a chronic condition, breast cancer. These two conditions were selected for study inclusion because WTP and SG are recommended for utility elicitation in both conditions but empirical evidence suggests that WTP may be more appropriate for acute conditions [12,14]. The acute condition, CINV, was selected. According to patients, CINV is regarded to be the most distressing symptom of cancer treatment [22–26]. CINV has been reported to occur in 60% to 100% of patients receiving chemotherapy [22]. Intensity, duration, and frequency of CINV vary, depending on the emetogenicity and prior patient experiences with chemotherapy [27,28]. Typically, a temporary symptom, CINV lasts 1 to 5 days with an average of 3 days after chemotherapy [27,28]. In this study, respondents were provided with background information describing cancer, chemotherapy, and its side effects including CINV and two therapies to prevent CINV: standard and optimal therapy [13]. The scenarios of standard and optimal therapy were based on the impact of partial and complete alleviation of CINV assumed to result from metoclopramide and ondansetron. All scenarios used in the study were pilot tested and reviewed by three clinical oncology practitioners; definitions of partial and complete alleviation of CINV are published in detail elsewhere [14].


INSTRUCTIONS: Now that you have used the willingness to pay method, chance board (computer slides) and thermometer (blue felt board) to rate choices in healthcare, we would like you to compare the different methods.

For example, suppose you were asked which of the three methods willingness to pay (W), chance board (C), and thermometer (T) required the most technology to perform. If you believe that T and W are of equal rating with C requiring a higher level of technology you may respond the following way..

	Most								Least
Technology use	C		T, W						

Using the following abbreviations willingness to pay (W), chance board (C), and thermometer (T) in your comparisons please compare the three methods for the following four questions.

	All decisions easy		Most decisions easy		Indifferent		Most decisions difficult		All decisions difficult
Difficulty of making decisions									

	Very Clear		Somewhat Clear		Indifferent		Somewhat Unclear		Very Unclear
Clarity of test									

	Very Reasonable		Somewhat Reasonable		Indifferent		Somewhat Unreasonable		Very Unreasonable
Reasonableness for decision making									

	Very Comfortable		Somewhat Comfortable		Indifferent		Somewhat Uncomfortable		Very Uncomfortable
Comfort for use in decision making									

Figure 6 GPS.

The chronic condition selected for study inclusion was breast cancer, regarded as one of the three most common cancers, the others being colorectal and lung cancer. In 1999, an estimated 43,300 women died of breast cancer in the United States. Breast cancer treatment serves as a paradigm for why quality of life research is important. Changes in breast cancer treatment over the past 20 years were based on patient quality of life given that different therapy options did not prolong life [29]. In this study, respondents were provided with background information describing breast cancer and its treatment. Therapies described breast cancer treatment and a hypothetical cure. The breast cancer scenarios used in this study described a situation 1 year after treatment; therefore, acute symptoms such as CINV were not expected to confound preference scores elicited.

Preference. Respondent preferences were elicited for six scenarios in each interview using VAS and SG techniques [5,6]. Scenarios included perfect health, death, their own health as defined by Furlong et al. [7], and three hypothetical cancer states [14,15]. In CINV interviews scenarios included complete, partial, and no alleviation of CINV, and breast cancer interviews included a hypothetical breast cancer cure, treatment, and "do nothing," in which there was no therapy for breast cancer recurrence scenarios.

VAS is a vertical, calibrated interval scale labeled with anchors 0 and 100 for the least and most desirable health states, respectively [7]. Respondents were asked to designate their most desirable and

least desirable health states as 0 and 100 using marked arrows (Fig. 4). The remaining four health states were assigned values by respondents in chronological order—from second to the fifth most preferred health state using the remaining four arrows.

A SG top-down titration method was used to elicit SG weights because of its greater efficiency and precision compared to the more commonly used ping-pong approach [30,31]. The SG method was presented graphically on computer using Microsoft PowerPoint 2000 (Version 2000, 1999, Microsoft Corporation, Redmond, WA). SG used same six health state descriptions presented in the ranking and feeling thermometer. The most and least desirable health states, as defined by the respondent in ranking were used as the anchor health states, which was similar to VAS. Respondents were presented with a series of four sets of slides, four intermediate health states descriptions, in chronological preference order from most to least desirable.

WTP assessed three potential benefits in health preferences using the bidding game method. In the CINV survey, respondent WTP was measured for complete, partial, and no alleviation of CINV, and for breast cancer survey, respondent WTP was assessed for cure, treatment, and do nothing scenarios for breast cancer treatment.

Analysis

Analysis involved a two-step process [32,33]. First, MANOVA was performed. Significant test results for MANOVA were based on *F* statistics, derived from Wilks' lambda. If the *F* statistic was significant, the second step required running a univariate analysis of variance (ANOVA) for each significant dependent variable. The univariate ANOVAs were three-way mixed factorial in design, that is, assessment × (condition × preferences), or $2 \times (2 \times 3)$. The discriminatory power of GPS versus TS was determined based on statistical significance of results. Fishers' LSD was used for all post hoc tests when omnibus *F* was significant. All statistical tests were performed using SPSS 10.1 statistical software package (Version 10.1, 2000, SPSS, Chicago, IL). Based on convention all tests assumed a α level of 0.05 to be significant.

Results

Of the 126 respondents interviewed, 7 were dropped from the study for the reasons provided below, resulting in the inclusion of 119 respondents in the study. One interview was discontinued because of apparent distress observed by the interviewer. The respondent's mother had recently died of breast cancer. Two respondents did not keep their scheduled second interviews. Four respondents were dropped because of lack of understanding in health utility measures used and belief in the hypothetical scenarios. In addition, one of these four did not appear to believe the stated purpose of the study and questioned its use for market research; therefore, responses had potential for strategic bias.

Thus, of the 252 interviews completed (i.e., 126 respondents each completing two interviews), 238 interviews (94.4%) (i.e., 119 respondents each fully completing 2 interviews) were completed included in study for analysis. Usable responses were collected from 66 TS and 53 GPS surveys. Data was collected from March 2000 to June 2000 at Ohio State University. Sociodemographic variables from respondents randomized to TS and GPS groups were similar indicating that differences observed were due to factors tested and not inherent differences between respondents assigned to one group or the other (Table 1). Mean age of respondents in this study was 29.4 years (SD 7.34). The majority of respondents were young, single, full-time graduate students with incomes less than or equal to US $20,000.

Means and standard errors of the mean (SEM) for TS and GPS groups are summarized for each of the dependent variables (difficulty, clarity, reasonableness, and clarity) for the three factors, condition, assessment, and preference, in Table 2. Overall Table 2 is notable for greater spread of means in GPS over TS and low mean ratings of less than five for all measures with clarity of text having the lowest mean rating and difficulty with decision making

Table 1 Sociodemographic characteristics of respondents randomly assigned to TS or GPS

Characteristic	TS ($n = 66$)	GPS ($n = 53$)
Age (years)		
Mean (SD)	29.4 (6.9)	29.3 (7.9)
Range	22–59	21–56
Marital status, *n* (%)		
Single	40 (60.6)	37 (69.8)
Married/cohabitating	26 (39.4)	16 (30.2)
Employment status (%)		
Student	54 (81.8)	43 (81.1)
Employed	12 (18.2)	10 (18.9)
Education		
High school or less	2 (3.0)	4 (7.5)
College	35 (53.0)	23 (43.4)
Graduate school	25 (37.9)	22 (41.5)
Missing	4 (6.1)	4 (7.5)
Income (%)		
≤$20,000	35 (53)	31 (58.5)
>20,000	31 (47)	22 (41.5)

Table 2 Summary of means (SEMs) for TS and GPS groups for the three factors: condition, assessment, and preference

	Assessment: TS ($n = 66$)/GPS ($n = 53$)					
	Preference			Preference		
Condition	VAS	SG	WTP	VAS	SG	WTP
Acute						
Difficulty	2.92 (0.23)	3.43 (0.27)	3.58 (0.27)	2.28 (0.20)	4.23 (0.35)	3.75 (0.26)
Clarity	1.59 (0.14)	1.75 (0.13)	1.56 (0.10)	1.42 (0.09)	1.81 (0.19)	1.87 (0.20)
Reasonableness	2.12 (0.20)	2.59 (0.21)	2.17 (0.21)	2.18 (0.19)	2.73 (0.26)	2.19 (0.21)
Comfort	2.33 (0.21)	3.15 (0.28)	2.77 (0.26)	1.97 (0.19)	3.43 (0.31)	3.19 (0.28)
Chronic						
Difficulty	2.72 (0.23)	3.86 (0.27)	3.68 (0.28)	3.15 (0.26)	4.71 (0.34)	4.53 (0.34)
Clarity	1.42 (0.10)	1.65 (0.12)	1.59 (0.11)	1.49 (0.15)	1.91 (0.19)	1.96 (0.19)
Reasonableness	2.01 (0.15)	2.35 (0.19)	2.56 (0.26)	2.51 (0.22)	3.02 (0.26)	3.36 (0.30)
Comfort	2.26 (0.18)	3.47 (0.29)	3.24 (0.30)	2.75 (0.28)	4.26 (0.35)	4.08 (0.37)

Note: Four dependent variables (difficulty, clarity, reasonableness, and clarity) were measured on a 9-point Likert scale with a lower score indicating greater preference. The midpoint was labeled as indifferent on all scales.

have the greatest mean rating in all cases; lower ratings were scored as being more preferable. Therefore, results showed that respondents perceived the text or required task to be clear for the three metric tested. However, as anticipated, respondents reported greater difficulty with SG and WTP methods compared to VAS, independent of assessment method used. For the means for VAS (2.8/2.6), SG (3.6/4.4), and WTP (3.6/4.0) for TS/GPS on a 9-point Likert scale, a lower score is rated as less difficult. Also, mean scores indicate that respondents found WTP a more reasonable metric in decision making for acute than for chronic conditions for both TS and GPS surveys. Respondents were less comfortable using SG than WTP for decision making in both acute and chronic conditions. The next section evaluates all differences for significance using MANOVA.

MANOVA results showed significant multivariate main effects ($P < .001$) for preferences (Wilks' lambda 0.81, $F_{8,110} = 9.290$) and condition (Wilks' lambda 0.56, $F_{4,114} = 6.375$), indicating that preference metric used (VAS, SG, and WTP) and condition assessed (acute or chronic) has an affect on respondents' perception of clarity of text, difficulty, reasonableness, and/or comfort in making decisions. Therefore, respondent attitude toward preference assessment was contingent on the type of condition (acute vs. chronic) and the preference method used. However, because these main effects have significant interactions, they cannot be considered in isolation.

MANOVA results showed significant multivariate interactions ($P < .05$) for condition × assessment (Wilks' lambda 0.89, $F_{4,114} = 3.421$) and condition × preferences (Wilks' lambda 0.87, $F_{8,110} = 2.087$), indicating that the type of condition (acute or chronic), assessment method (GPS or TS), and preferences metric (VAS, SG, WTP) has an effect on respondents' perception of clarity of text, difficulty, reasonableness, and/or comfort in making decisions, and this depends on the level of the three independent variables. To determine which of the dependent variables the significant results applied to, univariate ANOVAs were performed for each significant MANOVA main effects and interactions [32,34].

Univariate ANOVAs

Significant results for univariate ANOVAs were presented only for significant MANOVA and were presented for each dependent variable in turn. All dependent variables were scored on a 9-point Likert scale with a lower score being more desirable (minimum score = 1 to maximum score = 9). The next section presents univariate ANOVA for each significant dependent variable.

Clarity of text. For respondent perception of clarity of text main effect preferences was significant ($F_{2,234} = 5.660$, $P = .004$). Results showed respondents to perceive VAS (mean = 1.49) to be clearer than SG and WTP (means, 1.77 and 1.72, respectively) in terms of rating choices in health care. Of the four variables tested, clarity of text, difficulty of making decisions, reasonableness for decision making, and comfort in using for decision making, clarity of text had the best (lowest) rating (Table 2).

Difficulty making decisions. Results showed significant main effects for respondents perception of using VAS, SG, or WTP in making medical decisions, that is, preferences ($F_{2,234} = 27.04$, $P < .001$), and differences in ease of making decisions depending on the type of condition, that is, condition ($F_{1,117} = 8.68$, $P = .004$). Because there was a signif-

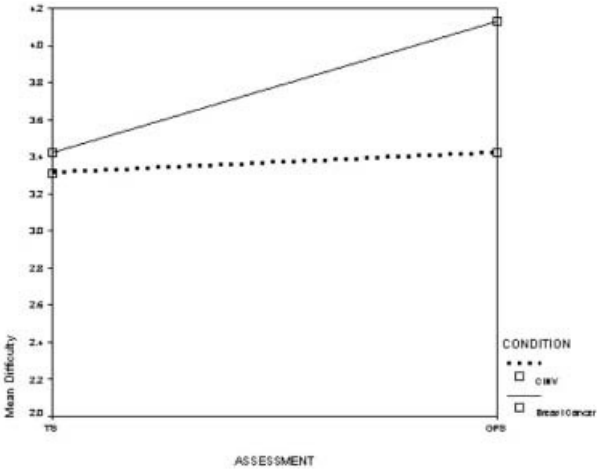


Figure 7 Interaction of condition × assessment ($F_{1,117} = 4.62$, $P < .05$).

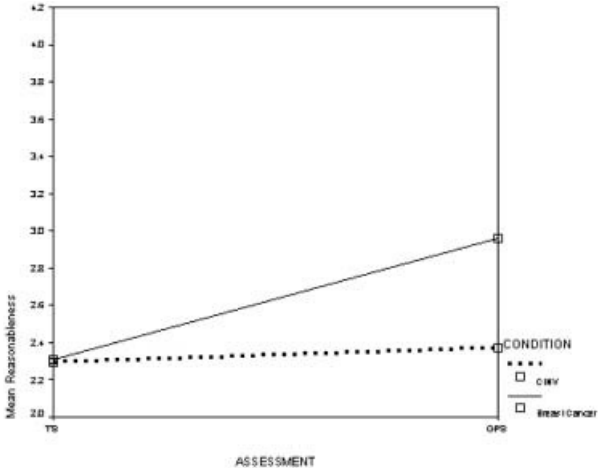


Figure 8 Interaction of condition × assessment ($F_{1,117} = 9.49$, $P < .05$).

icant interaction of condition × assessment ($F_{1,117} = 4.62$, $P = .034$), main effects cannot be interpreted in isolation (Fig. 7). The condition–assessment interaction shows that respondents perceived decision making in chronic conditions as more difficult than decision making in acute conditions. This result remained true independent of type of assessment technique used. However, the GPS technique resulted in a greater spread of means, indicating that respondents had more difficulty with decision making in chronic than in acute conditions. Post hoc tests showed that the difference was statistically different for GPS method. Also mean scores obtained for reasonableness when assessing chronic condition were statistically different for TS and GPS methods.

Reasonableness of measure for decision making. Main effects were significant for preferences ($F_{2,234} = 4.08$, $P = .081$) and condition ($F_{1,117} = 10.52$, $P = .002$). This implies that respondents perceived varying levels of the reasonableness of VAS, SG, and WTP methods for use in decision making. Furthermore, whether an acute or chronic condition was being valued affected respondent perception of reasonableness of the metric. Because there were two significant interactions, respondent perception of reasonableness of preference measure for use in decision making depends on the type of condition assessed (acute vs. chronic), the type of preference metric used, the in addition to method of assessment (Figs. 8 and 9).

The interaction condition × assessment ($F_{1,117} = 9.49$, $P = .003$) showed that the TS method resulted in respondents reporting utility assessment to be equally reasonable for acute and chronic conditions; however, utility assessment of chronic conditions was perceived as less reasonable (higher score) with GPS than with TS (Fig. 8). As indicated in Table 2, GPS scores showed greater spread in mean ratings. Post hoc test showed statistical significance for GPS scale. Also, mean scores obtained for reasonableness when assessing chronic condition were statistically different for TS and GPS methods.

The interaction condition × preferences ($F_{2,234} = 6.76$, $P = .001$) showed that VAS was reported as the most reasonable measure of preference (lower score) of the three tested, independent of type of

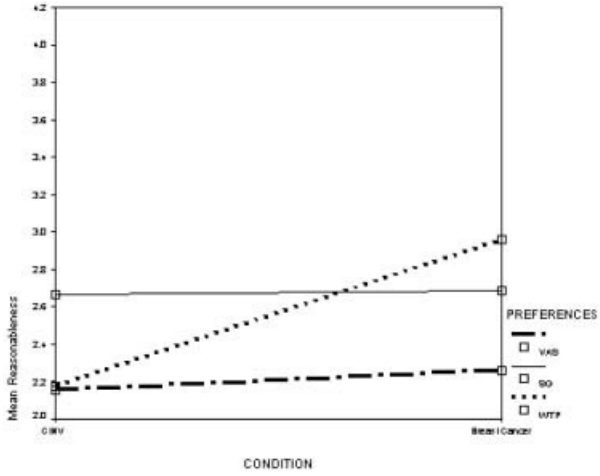


Figure 9 Interaction of condition × preferences ($F_{2,234} = 6.76$, $P < .05$).

condition. Post hoc tests showed that SG was equally reasonable in terms of mean scores for acute and chronic conditions tested; however, how VAS and WTP methods were perceived compared to SG varied significantly between acute and chronic conditions tested. SG was perceived as less reasonable (higher score) than VAS for acute and chronic conditions. Results showed that respondents reported WTP to be a more reasonable decision-making tool for acute than for chronic conditions, while VAS and SG were preferred by respondents when making decisions regarding chronic conditions (Fig. 9).

Comfort using utility measures in decision making. Testing respondent comfort with the preference measures in decision making showed significant main effects ($P < .001$) for preferences ($F_{2,234} = 24.44$) and condition ($F_{1,117} = 17.61$) and a significant interaction for condition × assessment ($F_{1,117} = 5.48$, $P = .021$) (Fig. 10). Significant main effects indicate that the choice of utility metric, VAS, SG, or WTP, impacts how comfortable a respondent is with making decisions in health care. Furthermore, comfort in health-care decision making can vary depending on the type of condition—whether it is acute or chronic. Respondents reported being more comfortable (lower score) assessing utility of temporary over chronic or rest-of-life conditions, independent of the assessment method (TS or GPS). However, post hoc tests showed this difference to be significant for GPS assessment method.

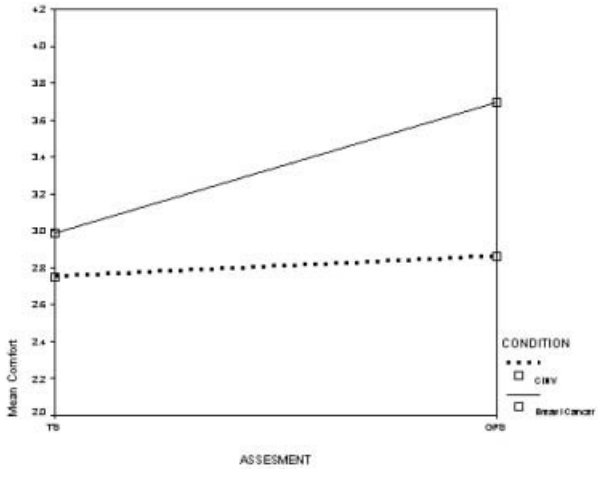


Figure 10 Interaction of condition × assessment ($F_{1,117} = 5.48$, $P < .05$.

Discussion and Conclusions

Minimizing survey length and increasing discriminatory power of data collection techniques are important goals of survey research in health care and marketing. Attainment of this goal has been reported in marketing literature [4]. This study evaluated the potential of GPS versus TS in assessing attitudes toward utility elicitation methods, VAS, SG, and WTP, assessment in health care. Our study findings, similar to results reported by Narayana [4] in the marketing literature, showed GPS to reduce survey length and assist respondents in being more discriminatory in their assessments. Greater discriminatory power of GPS over TS was shown in the greater spread of mean responses (Table 2) showing statistical significance using post hoc tests (Figs 7–10). An increased range of scores, but not necessarily standard deviations, resulted in statistically significant differences shown with GPS where none were noted for TS survey. This is readily seen in Figures 7, 8, and 10. The higher discriminatory power observed with GPS would make this method more desirable in attitude assessment. The greater spread and use of the Likert scale observed in this study is similar to that observed by Narayana. Narayana explained the heightened discriminatory power of GPS to be attributable to direct comparisons required with GPS versus indirect comparisons resulting with TS method. However, while results showed GPS to have higher discriminatory power in assessing respondent acceptability of preference measures, respondents appeared to find GPS more cognitively challenging in this study. When completing TS surveys, respondents did not refer back to previous answers when completing items, as per researcher observation. Respondents were not forbidden to do so. One explanation for this observation could be that because TS and GPS items were placed last in the survey, the respondents' hurry to complete the survey resulted only in the minimum task being performed. The GPS survey completion required respondents to review and compare all previous scores since it required direct comparisons—this was the minimal requirement. The respondents were not "cheating" using TS, just completing the minimal standard set. In summary, GPS required respondents to provide more thought in their comparisons between the utility measures. Therefore, GPS surveys did not take less time to complete, although GPS surveys were shorter in length. Thus, this study does not support conclusion of Narayana [4] that GPS results in reducing time of survey completion. Time of survey completion was not pre-

sented in this study because assessment method played a minor role in time to complete surveys as part of a larger study. Interested readers are referred to Franic [13].

The results of this study support contention of Torrance [16] of the value of VAS in utility assessment—the role of VAS in utility elicitation lies in familiarizing respondents with health-state scenarios. We also found VAS helpful in familiarizing respondents with health-state scenarios before administering SG and WTP methods (Fig. 9). Figure 9 shows that respondents found VAS to be a more reasonable measure in decision making. In this study, we found VAS to be valuable in easing respondents into preference elicitation methods. Of VAS, SG, and WTP methods, VAS was perceived as being the easiest to understand (i.e., clarity), and this helps to explain why VAS remains as one of the most popular methods for preference elicitation. Unlike SG, no probabilities were considered or chance of death, and unlike WTP, respondents do not have to consider trading money for health and were not limited by income in providing their preferences.

Although not the primary purpose of this study, results also showed that respondents regard WTP as a more reasonable decision-making tool when assessing acute interventions over SG and vice versa for chronic conditions. Furthermore, as part of a larger study requiring respondents to provide WTP bids for CINV and breast cancer states, bids in chronic conditions were notable for ceiling effects [14]. These conclusions are consistent with those reported in 2000 by Bala and Zarkin [12], suggesting that WTP may be a more appropriate utility measure than SG for acute conditions.

These study results indicate that assessing chronic conditions was more challenging than acute in decision making (Fig. 7). Respondents had more difficulty and less comfort when making decisions for chronic than for acute conditions. This was to be expected given that breast cancer has received much media attention of late and is potentially life-threatening, while CINV, although dreaded by chemotherapy patients, is not fatal and is short-lived, and thus this supports validity of results reported. These results also show that respondents regard WTP as a more reasonable decision-making tool when assessing acute interventions in preference to SG and vice versa for chronic conditions (Fig. 9). This is supported by study results of a larger study where respondents refused to gamble for short time frames (3 days) compared with rest-of-life scenarios with SG, which was consistent with a "chronic" format [14].

In summary, our results showed GPS could be used successfully in health outcomes surveys. Given the more complex nature of the task, we recommend the inclusion of an example to respondents on how to complete the task (Fig. 2). The GPS method compared to the TS method emphasizes differences perceived by respondents for the variables tested. The result was that GPS highlighted perceived respondent differences, enabling these differences to present themselves as statistically significant. GPS did successfully reduce survey length, while not compromising data quality.

This study was supported by a grant awarded by The Columbus Medical Research Foundation, Columbus, Ohio.

Presented at the ISPOR 4th Annual European Congress, Cannes, France, November 11–13, 2001.

References

1 Dillman D. Mail and Internet Surveys. (2nd ed.). New York: Wiley, 2000.
2 Fowler F. Improving Survey Questions: Design and Evaluation. (1st ed.). Thousand Oaks: Sage Publications, 1995.
3 Goldstein WM, Hogarth RM. Judgment and decision research: some historical context. In: Research in Judgment and Decision Making. New York: Cambridge University Press, 1997.
4 Narayana CL. Graphic positioning scale: an economical instrument for surveys. J Mark Res 1977; 14:118–22.
5 Torrance GW. Social preferences for health states: an empirical evaluation of three measurement techniques. Soc Econ Plan Sci 1976;1:129–36.
6 Torrance GW, Feeney D. Utilities and quality adjusted life years. Int J Technol Assess Health Care 1989;5:559–75.
7 Furlong W, Feeney D, Torrance GW, et al. Guide to Design and Development of Health State Utility Instrumentation. Center for Health Economics Policy Analysis Working Paper Series. Paper No. 90-9. Hamilton (Ontario): McMaster University, 1990.
8 Torrance GW. Measurement of health state utilities for economic appraisal. J Health Econ 1986;5: 1–30.
9 Torrance GW. Preferences for health states: a review of measurement methods. In: Clinical and Economic Evaluation of Perinatal and Developmental Medicine. No. 20. Evansville: Mead Johnson and Johnson, 1982.
10 Sackett DL, Torrance GW. The utility of different health states as perceived by the general public. J Chron Dis 1978;31:697–704.

11 Adams PF, Marano MA. Current estimates from the National Health Interview Survey, 1994. Vital Health Stat 1995;10:4.
12 Bala MV, Zarkin GA. Are QALYs an appropriate measure for valuing morbidity in acute diseases? Health Econ 2000;9:177–80.
13 Franic DM. Relating Quality Adjusted Life Years to Contingent Valuation: Acute Versus Chronic Conditions [dissertation]. Columbus: The Ohio State University, 2000.
14 Franic DM, Pathak DS, Gafni A. Are health states "timeless": a case study of an acute condition—post chemotherapy nausea and vomiting. J Eval Clin Prac 2003;9:69–82.
15 Franic DM, Pathak DS. Effect of including (versus excluding) fates worse than death in utility assessment. Int J Technol Assess Health Care 2003;19:346–60.
16 Torrance GW, Feeny D, Furlong W. Visual analogue scales: do they have a role in the measurement of preferences for health states? Med Decis Making 2001;21:329–34.
17 Gold MR, Siegel JE, Russell LB, Weinstein MC. Identifying and valuing outcomes. In: Cost-effectiveness in health and medicine. New York: Oxford University Press, 1996.
18 Spector PE. Quantitative Applications in the Social Sciences Series, #7. Research Designs. In: Sullivan JL, Niemi RG, eds., Beverly Hills: Sage Publications, 1987.
19 Flowers CR, Garber AM, Bergen MR, et al. Willingness to pay utility assessment. feasibility of use in normative patient decision support systems. In: Proceedings of the American Medical Informatics Association Fall Symposium 1997. Philadelphia: Hanley Belfus, 1997. p. 223–7.
20 Zielger JK, Hornberger JC, Pao V, et al. Assessing the Effects of Implantable Cardioverter Defibrillators on Quality of Life Using an Automated Computer Interview. Technical report, San Diego, University of California, August 2003.
21 Lenert LA, Cher DJ, Goldstein MK, et al. The effect of search procedures on utility assessment. Med Decis Making 1998;18:76–83.
22 Kwong WJ, Parasuraman TV. Cost-effectiveness analysis of oral ondansetron and prochlorperazine for preventing nausea and vomiting after moderately emetogenic chemotherapy. Pharm Pract Manag Q 1999;19:28–41.
23 Coates A, Abraham S, Kaye SB, et al. On the receiving end—patient perception of the side effects of cancer chemotherapy. Eur J Cancer Clin Oncol 1983;19:203–8.
24 Levine AM, Richardson JL, Marks G, et al. Compliance with oral drug therapy in patients with hematologic malignancy. J Clin Oncol 1987;5: 1469–76.
25 Lindley CM, Hirsch JD, O'Neill CV, et al. Quality of life consequences of chemotherapy induced emesis. Qual Life Res 1992;1:331–40.
26 Richardson JL, Marks G, Levine A. The influence of symptoms and side effects of treatment on compliance with therapy. J Clin Oncol 1988;6:1746– 52.
27 Osoba D, Zee B, Warr D, et al. Effect of postchemotherapy nausea and vomiting on health-related quality of life. Support Care Cancer 1997;5:307–13.
28 Grunberg SM, Boutin N, Ireland A, et al. Impact of nausea and vomiting on quality of life as a visual analogue scale-derived utility score. Support Care Cancer 1996;6:435–9.
29 Kuchler T. Quality of life and breast cancer—the learning curve never ends. Eur J Cancer 1998;34:277–8.
30 Clarke AE, Goldstein MK, Garber AM, et al. The effect of assessment method and respondent population on utilities elicited for Gaucher disease. Qual Life Res 1997;6:169–84.
31 Lenert L, Morss S, Goldstein M, et al. Measurement of the validity of utility elicitations. Med Care 1997;35:915–20.
32 Hair JF, Anderson RE, Tatham RL, Black WC. Multivariate Data Analysis with Readings. (3rd ed.). New York: MacMillan, 1992.
33 Stevens JP. Applied Multivariate Statistics for the Social Sciences. (4th ed.). Mahwah (NJ): Lawrence Erlbaum, 2001.
34 Keppel G. Design and Analysis: A Researcher's Handbook. (3rd ed.). Englewood Cliffs (NJ): Prentice Hall, 1991.