Telephone Interviews vs. Workstation Sessions for Acquiring Quality of Life Data

Michael W. Kattan, Ph.D.,[1] Paul A. Fearn, BA,[1] Scott B. Cantor, Ph.D.,[2] Jim Hu, M.D.,[3] Mark E. Cowen, M.D., SM,[4,5] R. Brian Giesler, Ph.D.,[6] Brian J. Miles, M.D.[3]

[1]Memorial Sloan-Kettering Cancer Center, New York, [2]Department of Medical Specialties, University of Texas M.D. Anderson Cancer Center, Houston, Texas [3]Scott Department of Urology, Baylor College of Medicine, Houston, Texas, [4]Department of Medicine, St. Joseph Mercy Hospital, Ann Arbor, Michigan [5]University of Michigan Medical School [6]Indiana University School of Nursing and School of Medicine, Indianapolis, Indiana

ABSTRACT

Patient quality of life data can be acquired in a variety of ways, including over the telephone and through computerized questionnaires. However, if the method of collection produces different results, medical decisions regarding appropriate and cost-effective care may be influenced by collection method. We conducted an experiment where subjects had two quality of life measures, the time trade-off and rating scale utilities, assessed both in telephone interivews and via computer touchscreens. The order of telephone and touchscreen was randomized. We found that rating scale utilities were similar whether obtained via the telephone or via touchscreen regardless of which was done first. However, patients who had their time trade-off utilities assessed over the telephone first did not provide as consistent responses as those elicited first via touchscreen (p=0.01). Caution is suggested when considering eliciting time trade-off over the telephone with subjects who have not had time trade-off elicited previously.

INTRODUCTION

The patient's evaluation of his or her current state of health, based on formal utility assessment methods, is integral to decision analyses and cost-effectiveness analyses. [1] As these analyses continue to grow in popularity, so will the demand for patient utilities. In conjunction with the demand for utilities in applied analyses, a substantial amount of basic research in utility assessment needs to be done. For the most part, it is often unclear why people have the utilities that they do. Not much is known about the stability of utilities over time, [2] and many health states do not yet have utility estimates. [3]

Moreover, few utility assessments have been done using a national sample, which is necessary for national health policy recommendations. [4] Furthermore, many decision analyses are sufficiently sensitive to utility estimates [5] such that individual utility assessments need to be done to obtain appropriate treatment recommendations at the individual patient level (*i.e.*, group means may not be sufficient). [6] All of these limitations of our knowledge of utilities drive the need for more utility assessment studies.

A major reason for the lack of utility assessment studies to answer these questions is that utility assessment is expensive and time-consuming. Utility assessment typically requires a trained facilitator who has props,[7] such as a computer with customized software.[8] Utility assessment also often requires in-person interaction between the subject and facilitator. This typically demands travel for the subject, facilitator, or possibly both. Thus, a less expensive utility assessment method would potentially accelerate progress in answering some of the questions surrounding utilities, provided it had validity comparable to the personal interview.

One way to facilitate utility assessment would be to have software that a telephone agent, who was not necessarily a utility assessment expert, could operate during a phone call. The software would control the iterative nature of utility assessment by providing the appropriate responses for the agent to use. In other words, the answers provided by the subject are input into the software which then provides the text for the telephone agent's next question. This approach eliminates most travel concerns, as well as training requirements for the facilitator, who has been replaced with a telephone agent with specialized utility assessment software.

The purpose of this study was to determine the association between utilities provided by subjects using a touchscreen and utilities elicited over the telephone by an untrained agent with scripted software. Our goal was, as much as possible, to isolate the effect of the utility assessment method by meas-

1091—8280/99/$5.00

uring the same subjects using both methods over a short period of time. Two utility assessment techniques, the time trade-off (TTO) and rating scale (RS), were of interest because they are generally felt to be easier for subjects to understand than other utility assessment techniques.[9]

METHODS

Every Friday, 50-70 men come to a free prostate cancer screening program at our university hospital. Men must register in advance, so the participants were known ahead of time. The evening before the free screening, half of the patients were randomly selected and called by a single interviewer. The interviewer read from a computerized script that elicited the TTO and RS utilities for the subjects' current state of health. These were essentially "cold calls" for the patients, as they had no visual aids sent to them in advance of the telephone call. The next day, as part of the screening process, all patients were directed to a human facilitator using a computer graphics display as a prop to illustrate the TTO and RS. Patient responses were entered into the computer by the facilitator, so that illiterate or vision-impaired patients were still able to participate. On Monday of the following week, the half of patients who were not called in advance were telephoned and read the standardized TTO and RS computer scripts to obtain their utility. For all utilities, patients evaluated their present health state, not the utility of a perceived health state. For both the telephone and touchscreen assessments, TTO was measured before RS. Both over the telephone and face to face, the bisecting approach of the TTO method was used as follows.

Men were first provided with their actuarial life expectancy based on age-specific U.S. mortality tables. Next, the subject was offered a choice between his own life with unknown life expectancy and guaranteed perfect health (followed by immediate guaranteed death) for a period of time equal to his actuarial life expectancy. If he chose his own life, his TTO was assigned the value 1.0, else he was presented with subsequent offers of perfect health. The "bisecting" approach was used, such that the subsequent offers were the midpoints between the maximum acceptable offer and maximum unacceptable offer of years of guaranteed perfect health. The TTO continued until the difference between upper and lower limits was less than 1 year. Figure 1 illustrates the touchscreen screens that were used as props for the facilitator. Figure 2 contains a screen capture of the computerized script that was used by the telephone agent.

Statistical Analysis

Analysis of covariance (ANCOVA) models were formed separately for TTO and RS utilities. These ANCOVAs included the telephone utility as a covariate and whether the subject was called before or after the utility assessment as an indicator variable. The response variable was touchscreen utility. Interactions between the indicator variable and covariate were tested, since significant interactions would indicate that interview order affected the relationship between the telephone and in-person utilities. Linear regressions were constructed and Pearson correlation coefficients were calculated to more closely examine the relationship between the telephone and in-person derived utilities.

Sample Size

We expected to find that utilities measured over the telephone would be highly correlated with utilities measured in person. Cohen[10] defines a large correlation as 0.5 or higher. We chose a sample size that would be large enough to provide adequate power (80%) to detect a large correlation if it were

Figure 1. Screenshots from the Workstation

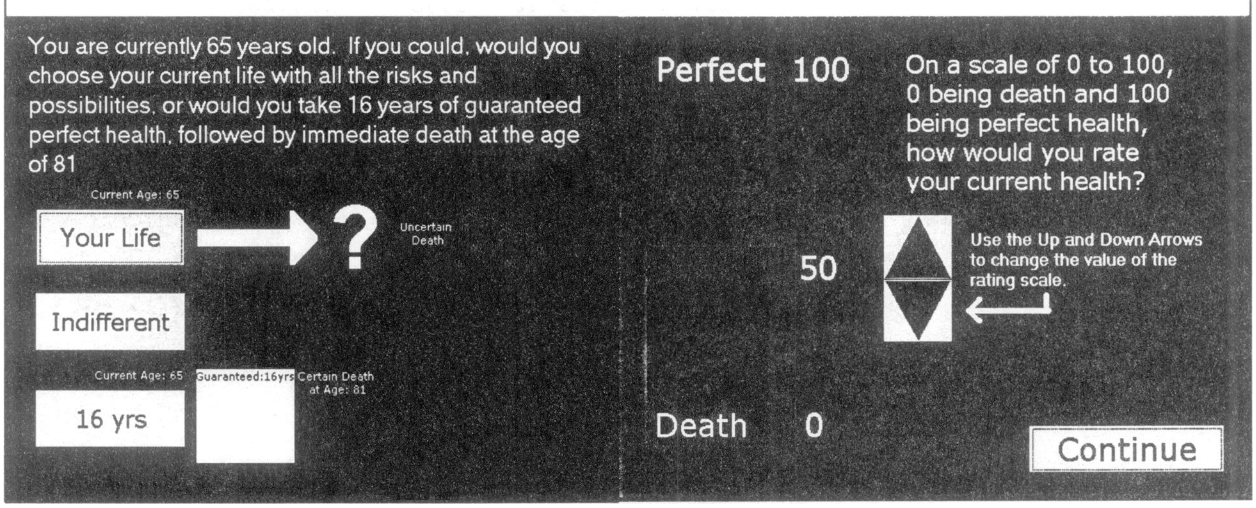
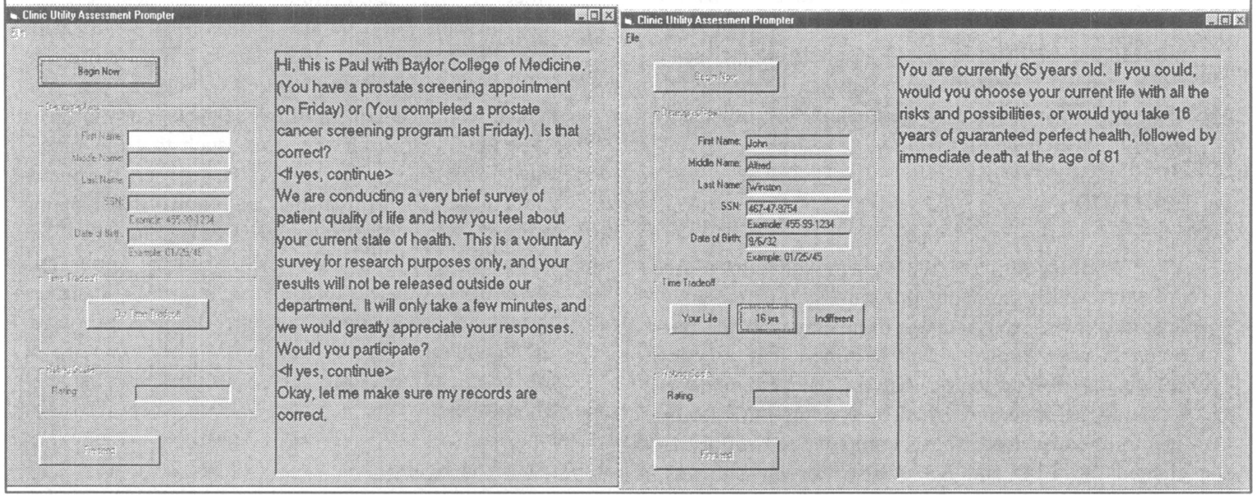


Figure 2. Sample Screens from the Computerized Script



present. In other words, we wanted to be confident that a strong correlation was unlikely if we did not find support for it in the sample collected. A sample size of 29 subjects provides 80% power to detect a Pearson correlation coefficient of 0.5 with a 5% risk of a type-I error.

RESULTS

A total of 64 men participated in the study. However, 19 subjects were not interviewed by telephone due to our inability to reach them. Analysis of covariance models for RS suggested no effect of whether the subject was telephoned prior to the in-person interview, either as a main effect (p=0.24) or an interaction (p=0.44) (see Table 1). Therefore, a reduced model was fit as a linear regression with the telephone RS as the only predictor of in-person RS (see Table 2). The intercept term was not significantly different from 0 (p=0.25), suggesting no bias to the telephone RS. The Pearson correlation coefficient between telephone and face-to-face RS utilities was 0.65 (p=0.01).

Analysis of covariance models for TTO detected an interaction between the order indicator variable and the telephone covariate (p=0.01), suggesting that the relationship between the telephone and touchscreen utilities depended upon whether the patient had the telephone interview prior to or after the touchscreen interview (see Table 3). Specifically, the intercept term for the ANCOVA model was not significant, but the main effect of the ordering of the utilities (i.e., telephone before or after) was significantly different from 0 (p=0.01). Thus, the models for the telephoned-first and telephoned-second groups appear to be different. The telephoned-first model

Table 1. Analysis of Variance Table for Rating Scale

	Df	Sum of Sq	Mean Sq	F Value	Pr(F)
PHONE-RS	1	0.5094428	0.5094428	30.84989	0.0000019
CALLED-FIRST	1	0.0237426	0.0237426	1.43776	0.2373861
INTERACTION	1	0.0098546	0.0098546	0.59676	0.4442484
Residuals	41	0.6770577	0.0165136		

Table 2. Linear Regression Results for Rating Scale

	Coefficients			
	Value	Std. Error	t value	Pr(>|t|)
(Intercept)	0.1411	0.1203	1.1733	0.2471
PHONE-RS	0.8237	0.1484	5.5520	0.0001

Table 3. Analysis of Variance Table for Time Trade-Off

	Df	Sum of Sq	Mean Sq	F Value	Pr(F)
PHONE-TTO	1	0.5445957	0.5445957	59.96278	0.0000001
CALLED-FIRST	1	0.0116545	0.0116545	1.28322	0.2638821
INTERACTION	1	0.0671033	0.0671033	7.38842	0.0095764
Residuals	41	0.3723714	0.0090822		

has an intercept of approximately 0.5 and a slope less than unity (Table 4). The telephone afterwards model has an intercept of 0 and slope very near unity, as would be expected under perfect agreement. Since the ANCOVA model was significant, this suggested that the relationship between telephone and touchscreen TTOs depends upon which is administered first. Therefore, correlations between telephone and touchscreen interviews were computed separately within the indicator variable for order of utility assessment. However, the correlation between the telephone and touchscreen TTO utilities was not dramatically stronger within the patients who were telephoned afterwards (r=0.80) versus the subjects who were telephoned in advance (r=0.78).

Table 4. Linear Regression Results for Time Trade-Off

	Coefficients			
	Value	Std. Error	t value	Pr(>|t|)
(Intercept)	-0.0364	0.1472	-0.2476	0.8057
PHONE-TTO	1.0111	0.1531	6.6059	0.0001
CALLED-FIRST	0.5000	0.1742	2.8703	0.0065
INTERACTION	-0.5007	0.1842	-2.7182	0.0096

DISCUSSION

This study was an experiment designed to evaluate whether utilities could be accurately measured over the telephone by a telephone agent who is not trained to elicit utilities but has a software script. Two utility techniques were evaluated, the rating scale (RS) and the time trade-off (TTO). The RS utilities as measured over the telephone were, in general, very similar to those obtained during in-person interaction with a trained facilitator. The association measure was quite strong (r=0.65, p=0.001), and it did not seem to make much difference whether the subject had experience with the rating scale. These results suggest it is relatively safe to administer a rating scale over the telephone to a new subject by someone with very little training in utility assessment, provided the interviewer has a canned software script.

However, the TTO may not be as reliable to administer over the telephone. The association between TTO utilities over the telephone and TTO utilities in an in-person encounter appear to be different (p=0.01) depending upon which is administered first. In other words, a cold call which administers the TTO appears to produce utilities which are different from those provided by the subject in a traditional utility assessment interview. In contrast, the subject who has had his TTO assessed in an in-person manner seems to provide a very similar TTO in a subsequent telephone interview.

There are important limitations to our study. First, our telephone response rate was 70%. We were unable to contact 19 of our 64 subjects over the telephone, resulting in data from 45 subjects for analysis. While 45 subjects may be larger than our minimum sample size calculations, they still represent a relatively small sample. More complex relationships between telephone and in-person utilities (*e.g.*, algebraic transformations such as a power function [11]) may be possible with a larger sample. Third, it may be possible to improve upon our computerized script. Our study cannot be considered a condemnation of all computerized utility assessment scripts for the time trade-off. And fourth, it is plausible that multiple telephone calls eventually result in better understood utilities, such that subsequent telephone administrations (without the initial face-to-face interview) produce reliable estimates as subjects eventually comprehend the exercise.

Another issue associated with our experimental design concerns the delay between utility assessments. Subjects may have tried to provide the same utility the second time as they did the first time. For example, the patient who has telephoned prior to the screening visit yielded a particular utility value, and he may have tried to yield the same value the next day thinking he was supposed to be consistent. However, such a phenomenon seems unlikely to explain our results. The reason for this is that patients who were called first participated in face-to-face assessment the next day, while patients who had in-person utilities elicited first (on Friday) were not telephoned until Monday. Thus, more time would have passed in the telephoned-second group, yet these were the subjects with better agreement.

Another explanation for the discrepancy between telephone and in-person TTO could be that the men's health state was more likely to change between assessments in the telephoned-first group. However, with regard to our example of prostate cancer test results, this explanation is improbable. None of the patients had their blood test results prior to their second utility assessment. The digital rectal exam was performed after the in-person utility assessment, so it should not have affected the relationship between utilities for the patients who were telephoned first. However, it would have the potential to affect the patients who were telephoned second since the exam occurs between utility assessments. However, this is the group with stronger association, making the exam result an unlikely factor in the utility relationship analysis.

In conclusion, our results suggest caution for those considering large scale TTO administrations over the telephone to randomly selected subjects. While such utilities are much needed both for applied and basic purposes, they may not be reliable if the subject has not had TTO elicited in an in-person manner previously. Further research is needed on utility assessment methods that maintain the reliability and validity of in-person interviews, but also can be administered quickly and at low cost.

REFERENCES

1. Gold MR, Patrick DL, Torrance GW et al. Identifying and valuing outcomes. In Gold MR, Siegel JE, Russell LB, Weinstein MC (eds). Cost-Effectiveness in Health and Medicine, New York: Oxford University Press, 1996.

2. Llewellyn-Thomas HA, Sutherland HJ, Thiel EC. Do patients' evaluations of a future health state change when they actually enter that state? Med Care. 1993; 31:1002-1012.

3. Torrance GW, Feeny D. Utilities and quality-adjusted life years. Intl J Technology Assessment in Health Care. 1989;5:559-575.

4. Kaplan RM. The hippocratic predicament: affordability, access, and accountability in American medicine. San Diego: Academic Press, 1993.

5. Kassirer JP. Incorporating patients' preferences into medical decisions. New Engl J Med. 1994;330:1895-1896.

6. Cowen ME, Miles BJ, Cahill DF, Giesler RB, Beck JR, Kattan MW. The danger of applying group-level decision analyses for the treatment of localized prostate cancer to the individual. Med Decis Making. 1998;18(4):376-380.

7. Furlong W, Feeny D, Torrance GW, Barr R, Horsman J. Guide to design and development of health-state utility instrumentation. CHEPA Working Paper Series #90-9, June 1990.

8. Clarke AE, Godstein MK, Michelson D, Garber AM, Lenert LA. The effect of assessment method and respondent population on utilities elicited for Gaucher disease. Quality of Life Research. 1997;6(2):169-84.

9. Torrance GW. Measurement of health state utilities for economic appraisal: a review. J Health Econ. 1986;5:1-30.

10. Cohen,J. Statistical Power Analysis for the Behavioral Sciences. Hillsdale, NJ: Lawrence Erlbaum Associates, 1988.

11. O'Leary JF, Fairclough DL, Jankowski MK, Weeks JC. Comparison of time-tradeoff utilities and rating scale values of cancer patients and their relatives: evidence for a possible plateau relationship. Med Decis Making. 1995;15:132-137.