

# An alternative approach for eliciting willingness-to-pay: A randomized Internet trial

Laura J. Damschroder<sup>\*1,3</sup>, Peter A. Ubel<sup>1,2,3,4</sup>, Jason Riis<sup>5</sup>, and Dylan M. Smith<sup>1,2,3</sup>

<sup>1</sup>HSR&D Ann Arbor Center of Excellence, Department of Veterans Affairs, Ann Arbor, MI

<sup>2</sup>Division of General Internal Medicine, University of Michigan

<sup>3</sup>The Center for Behavioral and Decision Sciences in Medicine, University of Michigan

<sup>4</sup>Department of Psychology, University of Michigan

<sup>5</sup>Department of Marketing, Stern School of Business, New York University

## Abstract

Open-ended methods that elicit willingness-to-pay (WTP) in terms of absolute dollars often result in high rates of questionable and highly skewed responses, insensitivity to changes in health state, and raise an ethical issue related to its association with personal income. We conducted a 2x2 randomized trial over the Internet to test 4 WTP formats: 1) WTP in dollars; 2) WTP as a percentage of financial resources; 3) WTP in terms of monthly payments; and 4) WTP as a single lump-sum amount. WTP as a percentage of financial resources generated fewer questionable values, had better distribution properties, greater sensitivity to severity of health states, and was not associated with income. WTP elicited on a monthly basis also showed promise.

Keywords: health, contingent valuation, willingness-to-pay, computerized elicitation, income.

## 1 Introduction

Many economists elicit people's willingness to pay (WTP) for healthcare interventions through contingent valuation surveys so that the benefits of those interventions can be valued in monetary terms (Diener, O'Brien, & Gafni, 1998; Klose, 1999; Olsen & Smith, 2001; Smith, 2003). This is despite many known biases that occur when attempting to elicit a dollar value from people for a good that is not usually directly available in the market; e.g., perfect health (Baron, 1997). Much literature focuses on developing consensus on the most valid method for eliciting WTP; putting aside any philosoph-

ical issues that question the validity of eliciting WTP through a single elicitation. Early WTP surveys elicited values using an open-ended question from a self-interest perspective to obtain personal use values; e.g. "how much would you be willing to pay to be cured?" (Smith & Richardson, 2005). These open-ended formats ask for WTP values without presenting a starting point value and without using a search routine to help respondents determine a value. Respondents are simply asked to give a dollar value. However, researchers have questioned the validity of this format because responses are prone to a high number of non-response or zero values and because responses are heavily skewed toward high values, perhaps, in part, due to strategic bias (Donaldson, Thomas, & Torgerson, 1997; O'Brien & Gafni, 1996). In response to these concerns, a U.S. Federal panel in 1993, led by Kenneth Arrow, concluded that "both experience and logic suggest that responses to open-ended questions will be erratic and biased" (Arrow et al., 1993, p. 4613).

Since then, researchers have moved away from eliciting WTP using an open-ended format and developed three types of closed-ended formats in an attempt to overcome shortcomings of the open-ended format. These "close-ended" formats ask respondents to say yes or no to a series of questions or to select a value from a pre-specified list. All three methods have methodological is-

\*The authors would like to thank Richard Smith for his insightful comments on earlier drafts of this paper. Also, thanks to Todd Roberts and Jennifer Heckendorn who helped administer and implement the survey. Financial disclosure: This research was supported by HSR&D Ann Arbor Center of Excellence, Department of Veterans Affairs and the National Institute on Child Health and Human Development Grant #R01HD040789. The funding agreement ensured the authors' independence in designing the study, interpreting the data, writing and publishing the report. The following authors are employed by the VA Ann Arbor Healthcare System: Laura J. Damschroder, Dylan Smith, and Peter A. Ubel. Dylan Smith is supported by a career development award from the Department of Veterans Affairs.

Direct Correspondence to: Laura J. Damschroder, University of Michigan Health System, 300 North Ingalls, Room 7C27, Ann Arbor, MI 48109-0429. Email: Laura.Damschroder@va.gov

sues, however. The bidding game is prone to starting-point bias (WTP changes depending on the starting value used to begin the bidding) and the payment card method is prone to range bias (WTP changes depending on the range of values presented) (Klose, 1999; Smith, 2000; Venkatachalam, 2004; Whynes, Wolstenholme, & Frew, 2004). The single-bounded discrete choice format is statistically inefficient and studies using this approach are very expensive to conduct because, all else being equal, it requires a larger sample size and more sophisticated design and analysis techniques (Smith, 2000; Venkatachalam, 2004). In addition, this format is prone to several biases including “yea-saying” where respondents have a tendency to agree with the amount presented (Yeung, Smith, Ho, Johnston, & Leung, 2006). A double-bounded choice format was derived to increase statistical efficiency. However, even responses from people who report a high level of certainty about their willingness to pay exhibit significant anomalies that increase as uncertainty increases (Watson & Ryan, 2006).

We believe the open-ended format deserves further exploration. Despite the strong statement we quoted earlier against using it, some researchers do not agree with the call to abandon the open-ended format (Smith, 2000). Although different formats produce different responses, it is not clear which format is superior (Venkatachalam, 2004). A recent study comparing alternate elicitation formats concluded, “. . . it would seem that the most informative elicitation format in the present context . . . appear[s] to be the open-ended format. . . [though this] format is nowadays distinctly unfashionable in health economics, having long since given way to supposedly-superior elicitation formats” (Whynes, Frew, & Wolstenholme, 2005, p. 384). Advantages of the open-ended format are that it does not introduce range or starting-point biases and it can be highly statistically efficient compared to discrete choice formats.

The open-ended format also has several clear disadvantages, however. This format may place a heavy cognitive demand on respondents. In fact, the other formats were developed, in part, to make the elicitation simpler and more realistic for respondents (Donaldson et al., 1997; Smith, 2000). Furthermore, asking for WTP in terms of dollars using an open-ended format requires using an unbounded response scale (a scale that starts at zero but with no defined upper end) that naturally contributes to the highly variable and skewed responses typically seen with open-ended WTP elicitations (Kahneman, Ritov, & Schkade, 1999). In addition, people may be more likely to give “strategic” values with an unbounded scale; a respondent may believe that the treatment has high intrinsic or social value and thus places a very high value not grounded in the reality of actually paying such a figure in the form of taxes or as an out-of-pocket expense (Arrow

et al., 1993). Conversely, a respondent may give an artificially low response in an attempt to influence the actual price eventually charged.

It could be that a more constrained, but still essentially open-ended approach might avoid some of the problems reviewed above. Specifically, eliciting WTP as a percentage of financial resources has two potential advantages. First, a percentage measure will force the use of a bounded 0–100 response scale creating a more statistically efficient scale measure (Kahneman et al., 1999). Generally, people are unable to map their preference for a health effect using a scale consisting of dollars with that starts at zero but with no clear maximum amount (an unbounded scale) (Payne, Bettman, & Schkade, 1999). Second, percentages involve smaller numbers (a 0–100 scale for the percentage formats versus 0 to an undefined maximum for the dollar formats) and people process smaller whole numbers more reliably. In one study, Thompson, Read, and Liang (1984) found that a percentage measure exhibited more significant associations with key independent variables such as the number of symptoms suffered by respondents and medications taken than did WTP expressed in dollars.

The purpose of the current study was to compare WTP values elicited as a percentage of financial resources to values elicited as dollars using open-ended formats. We predicted that the percentage method would be less prone to inconsistent responses, would be more sensitive to differences in severity across health states, and would show more desirable distributional properties. We asked for percentages based on “financial resources” rather than income because it is realistic to expect that many people would consider savings, borrowing power, and other financial resources to pay for a cure of a condition they want to avoid. Thinking about paying out amounts on a monthly basis rather than a single lump sum enables respondents to think of smaller quantities and the amounts proposed are likely to be more salient because many people budget their finances on a monthly basis. Advantages of the percentage format could be reduced or eliminated when monthly payments rather than lump sum payments are considered. Thus, we also introduced a second dimension against which to compare elicitation formats: a monthly timeframe versus a single lump sum amount.

The current study extends the studies done by Thompson and colleagues (the largest study, to date, that has elicited WTP as a percentage) in several ways. First, we introduce a within-subjects measure of sensitivity. Second we compare the effects of using a monthly timeframe to elicit WTP to a single lump-sum amount. Third, we focus specifically on distributional properties of responses to further assess percentage formats as a more efficient measure. Finally, the current study utilizes a larger sample, and surveys the general public instead of patients.

Table 1: WTP elicitation formats.

WTP Units:	Time Period:	
	Total	Monthly
\$US	Please type the <b>maximum dollar amount</b> you think you would be <b>willing and able</b> to pay for this treatment.  \$_____ (please enter only one amount)	Please type the <b>maximum dollar amount</b> you think you would be <b>willing and able</b> to pay <b>per month</b> for this treatment.  \$_____ <b>per month.</b> (please enter only one amount)
% financial resources	Please type the <b>maximum percentage of your financial resources</b> you think you would be <b>willing and able</b> to pay for this treatment.  _____ % of my financial resources. (please enter a number between 0 and 100)	Please type the <b>maximum percentage of your financial resources</b> you think you would be <b>willing and able</b> to pay <b>per month</b> for this treatment.  _____% of my financial resources <b>per month.</b> (please enter a number between 0 and 100)

## 2 Method

We elicited people’s WTP for curing two health conditions using a web-based survey over the Internet. We recruited respondents via an email sent to a sample of members in an Internet panel maintained by Survey Sampling International (SSI). This panel is made up of more than 1 million unique member households, recruited via random digit dialing, banner ads, and other opt-in techniques. Our study sample was stratified to mirror the U.S. census population based on age, gender, race, education level, and income. Upon completion of the survey, participants were entered into a drawing for cash prizes that totaled \$10,000.

### 2.1 Health state descriptions

We presented descriptions of two health states to each respondent: 1) a below-the-knee amputation (BKA) that moderately affects physical mobility; and 2) paraplegia, which significantly affects mobility. Detailed health state descriptions are in the appendix. We counterbalanced the order of the BKA and paraplegia health states.

### 2.2 WTP elicitation formats

We elicited each respondent’s WTP for a medical treatment that would permanently restore full physical functioning for each of the two health states. Respondents were randomly assigned to one of four elicitation formats, using a full-factorial two-by-two experimental de-

sign. We elicited WTP using one of two different units of measure (percentage of financial resources or dollars) and one of two different timeframes (on a monthly basis or an overall total). No durations for payments were specified. We chose percentage of “financial resources” instead of income for reasons already cited. Financial resources will typically be equal to or greater than income; thus, the underlying scale could represent values greater than income. The four versions (2 WTP measures X 2 timeframes), along with the specific questions we posed are presented in Table 1.

For each format, we first presented the description of the health state (listed in the appendix) and then asked the respondent to type in their response. The precise wording asking for a WTP amount depended on the format to which the respondent was assigned, as presented in Table 1. We then told respondents, “In answering this question, take into consideration the actual financial resources you have. We recognize that giving an exact amount may be difficult; just give the best estimate you can.” Our purpose with this instruction was to emphasize personal financial constraints before respondents gave a WTP amount. We elicited WTP for both health states from each respondent.

### 2.3 Outcome criteria and analysis approach

Analyses were performed using the native units and timeframe with which WTP was elicited; e.g., in terms of monthly percentage of financial resources. Our primary

study question was whether WTP expressed as a percentage of financial resources would result in higher quality responses and better distributional properties compared to WTP expressed in absolute dollars, and thus would show greater ability to detect differences between health states of different severity. We also wanted to explore whether WTP expressed on a monthly basis would improve properties of WTP responses and perhaps reduce any advantages observed of the percentage format.

We compared the four elicitation formats using five criteria:

First, we wanted to reduce the number of questionable WTP responses. Questionable WTP responses include missing values, values of zero, or WTP values that are the same for both health states. We used  $\chi^2$  tests to compare differences in frequencies for these types of occurrences between the formats. Those who gave missing or zero values for both health states were excluded from the remaining analyses.

Second, we assessed normality of WTP values in terms of skewness and kurtosis. Parametric models are often used to predict WTP responses and assume that WTP values and error terms are normally distributed. Even a small misspecification of the functional form in these analyses can result in large differences in predictions (Yeung et al., 2006).

Third, we assessed internal consistency with a simple ordinal consistency check. WTP values should reflect the lower impact that BKA has on mobility compared to paraplegia. Accordingly, we expect respondents' WTP for treatments to be lower for BKA compared to paraplegia. We excluded cases where the value was the same for both health states from this portion of the analysis and they were not included in the denominator. We used  $\chi^2$  tests to compare differences in the proportion of those who were ordinally consistent between the groups.

Fourth, we tested the sensitivity of each of the WTP elicitation versions for detecting differences between the two health states by computing Cohen's *d*-statistic as a measure of effect size (Cohen, 1988). Larger effect sizes indicate greater sensitivity and thus will require smaller samples to detect statistical differences between two health states.

Our final assessment was investigating the degree to which WTP values correlate with reported income for each of the four formats, using the Spearman rank correlation coefficient. Confidence intervals were computed using the bias-correction and accelerated bootstrap estimation method (Haukoos & Lewis, 2005). Two smaller-scale studies that elicited WTP as a percentage of wealth did not find this measure to be significantly associated with personal income (Schiffner et al., 2003; Thompson, 1986). Nonetheless, it is possible that an association would still persist in our study because people with low

incomes may have fewer discretionary finances available, even when expressed as a percentage (Donaldson, Birch, & Gafni, 2002). Though we did not have a prediction about whether WTP and income would be significantly associated with the WTP elicited using the two percentage formats, we did hypothesize that WTP as a percentage of wealth would have a lower association with income compared to WTP elicited as dollars.

### 3 Results

Compared to WTP expressed as absolute dollars, WTP expressed as a percentage of financial resources generates more usable values, greater sensitivity to differences in severity between health states, better distribution properties, and is not associated with income. Furthermore, asking WTP in terms of monthly amounts also shows promise.

#### 3.1 Respondents

Eight percent of those invited responded by clicking onto our survey using a link from within the email invitation. Of those who clicked onto the site, 75% ( $n=982$ ) completed the survey. Of those who completed the survey, 98% were included in the analyses, except where noted. 5 were excluded because they were under 18 years old, 15 said they intentionally gave wrong answers, and one gave invalid values (38,117 for both health states using the monthly percentage format). The rate of exclusions were similar across the four versions of survey ( $p=.22$ ). The remaining 961 respondents gave 1,812 non-zero and non-missing WTP valuations; 55 (6%) gave missing or zero WTP values for both health states.

The 961 respondents included in the analyses were not statistically different across the experimental groups with respect to demographic factors ( $p$ -values  $> 0.15$ ). Overall, 31% of respondents identified themselves as being a non-white race or Hispanic ethnicity. Self-reported mean age was 46 years ( $s.d.=16$ ). Median education was some college but no degree. Overall, 59% of respondents were women. Just under half (44%) of respondents identified themselves as having "average" economic status and 47% of respondents reported an income of \$40,000 or less.

#### 3.2 Questionable Values

55 (6%) respondents gave zero or missing values for both health states. Another 39 (4%) gave a zero or missing value for one health state. The rate of zero or missing values was comparable across the four versions (Chi-square;  $p=.60$ ). However, the rate of those who gave zero or missing values for both health states varied by

Table 2: Summary of outcome criteria.

	% Total n=208	% Monthly n=246	\$ Total n=243	\$ Monthly n=209
% respondents with the same WTP for both health states				
	35%	44%	43%	55% **
Skewness				
BKA	0.81 **	1.24 **	7.64 **	2.93 **
Paraplegia	0.07	0.72 **	5.42 **	2.00 **
Kurtosis				
BKA	2.74	3.97 *	70.77 **	15.19 **
Paraplegia	1.82 **	2.61	38.68 **	8.77 **
Spearman rank correlation coefficients for WTP and income				
BKA	0.01	0.07	0.30 **	0.33 **
Paraplegia	0.12	0.14 *	0.30 **	0.39 **

\*\* p<0.01, \* p<0.05

income (Wilcoxin rank-sum,  $p<.001$ ); three-quarters of these cases had income less than the median. It is possible that these subjects did not have any discretionary financial resources with which to pay for a cure (Smith, 2005). Respondents who gave zero or missing values for both health states were dropped from the remainder of the analyses.

Another type of potentially questionable value came from respondents who gave the same non-zero, non-missing value for both health states. Table 2 shows the distribution of these cases. Participants assigned to a monthly format (dollar or percentage) gave the same WTP for both health states more often than those who were not ( $p=0.004$ ). Participants assigned to a percentage format (monthly or lump sum) gave the same WTP values for both health states less often than those who were not ( $p=0.008$ ). The combined effect resulted in only 35% of participants who were assigned to the total percentage format giving the same WTP for both health states while over half (55%) of participants assigned to the monthly dollar format did so ( $p<0.001$ ).

### 3.3 WTP values

Table 3 shows mean and median WTP values for each of the elicitation formats. Respondents were willing to pay \$30,276 in total or \$252 per month to cure BKA when WTP was elicited as dollars. WTP in terms of percentages were 35% of financial resources as a total amount and 28% when elicited on a monthly basis. To cure paraplegia, respondents were willing to pay \$73,968 in total

or \$325 per month; WTP, when elicited as percentages was 53% as a total amount and 39% on a monthly basis.

### 3.4 Ordinal consistency of responses

On average, 88% of respondents who gave different WTP values for the 2 health states were willing to pay more to cure paraplegia than for BKA (Table 3). The rate of ordinal consistency did not vary by whether or not WTP was elicited by month ( $p=0.41$ ). However, respondents assigned to a percentage format had a higher rate of ordinal consistency (91%) compared to those assigned to a dollar format (84%) ( $p=0.03$ ).

### 3.5 Sensitivity to differences in severity

WTP means for the two health states were significantly different, regardless of the elicitation format ( $p$ -values $<0.001$ ). However, the differences in effect size across the versions varied considerably. The percentage format on a total basis had nearly a 3 times larger effect size than the corresponding dollar format. The effect size for the percentage format on a monthly basis was over 1.5 times larger compared to the effect size for dollars elicited on a monthly basis. As seen in Table 3, these differences in effect sizes translate to dramatic differences in sample sizes needed to detect differences between the two health states.

Table 3: WTP values by version

		WTP elicited as:			
		% Total	% Monthly	\$ Total	\$ Monthly
BKA <sup>1</sup>	Mean	35	28	30,276	252
	Median	25	20	8,500	150
	(s.d.)	(28)	(23)	(91,947)	(289)
Paraplegia	Mean	53	39	73,968	325
	Median	50	30	10,000	200
	(s.d.)	(30)	(27)	(209,814)	(324)
% respondents willing to pay more to cure paraplegia than to cure BKA <sup>1</sup>		93%	88%	84%	84%
Cohen's d effect size <sup>2</sup>		0.70	0.57	0.24	0.35
Sample size required <sup>3</sup>		16	25	137	64

1. Below-the-knee amputation. Include only respondents who gave different WTP values for the two health states.
2. Effect size, used in power analyses, for comparing difference in mean WTP for BKA and paraplegia for each of the elicitation versions.
3. Sample size that would be needed to detect the difference in mean WTP with 80% power and 5% alpha level for each of the elicitation versions.

### 3.6 Normality of responses

As can be seen in Table 3, there is a wide disparity between mean and median values, especially for the dollar amount formats, indicating highly skewed distributions. Indeed, Table 2 shows that the skew statistics for the dollar value formats were 2.0 or higher, indicating a distribution that is skewed toward high positive values. The skew statistics for 3 out of 4 of aggregate values using percentage formats were less than 1.0. However, the only distribution of responses that was statistically similar to a normal distribution were WTP values elicited in terms of the total percentage of financial resources for curing paraplegia ( $p=.7$ ). Most response distributions exhibited significant kurtosis, with kurtosis statistics as high as 71 for WTP values expressed as dollars. A normally distributed set of responses would have a statistic equal to 3.0. WTPs in terms of percent of financial resources are much closer to this target value and in fact, 2 of the 4 sets of responses are statistically similar to that expected for a normal distribution ( $p\text{-values}>0.2$ ).

### 3.7 Correlation with income

WTP expressed as absolute dollars, in monthly and total timeframes, were both significantly correlated with income for below-the-knee amputation and paraplegia. These correlations were all significantly higher than cor-

relations obtained by using the percentage formats ( $p\text{-values}<.01$ ), except that the lump sum dollar format was only marginally higher than using the monthly percentage format when eliciting values for curing paraplegia ( $p=.06$ ). WTP expressed in terms of percentage of financial resources was significantly correlated with income only for paraplegia and only if expressed on a monthly basis.

## 4 Discussion

Asking people to give their WTP as a percentage of financial resources instead of asking for WTP as dollars is a promising way to improve WTP measures that are typically plagued by undesirable properties. We also evaluated timeframe and found that the advantages of the percentage format persisted when a "per month" instead of a lump sum method was used. The percentage lump sum format yielded the fewest respondents who gave the same value for two different health states with clearly different levels of severity and yielded the highest rate of respondents who were ordinally consistent (WTP was higher for curing the health state with the more severe impairment [paraplegia] than for the less severe physical impairment [BKA]). The two percentage formats were substantially more sensitive to differences between health states and

thus more statistically efficient compared to WTP expressed as absolute dollars in total or on a monthly basis. This improvement in sensitivity translates to an 8-fold reduction in the sample size required to detect comparable differences in other studies when comparing the best performing format (WTP as a total percent of financial resources) to the worst performer (WTP as total dollars). Both percentage formats yielded more nearly normally distributed WTP values compared to WTP in either monthly or total dollars. The worst performer on every criterion was WTP expressed as absolute dollars; either monthly or total, depending on the criteria. The superior psychometric properties assessed in this study for WTP measured as a percent are good news considering that though many researchers recognize the challenging distribution properties of WTP values used in CBAs (cost-benefit analyses), there has been little consensus on what to do about it (Donaldson, 1999).

On average, participants were willing to pay 28% of their financial resources on a monthly basis (35% on a total percentage basis) to cure BKA and 39% (53% on a total percentage basis) to cure paraplegia in our study. The percentage for curing BKA is higher than the 17% (Thompson, Read, & Liang, 1984) and 22% (Thompson, 1986) for relief of arthritis symptoms in the studies by Thompson. Schiffner and colleagues also elicited WTP directly as a proportion of monthly income. Pre-treatment, psoriasis patients were willing to pay 14% of their income for a cure (Schiffner et al., 2003). It is difficult to assess whether the values obtained in our study are out of line with these previous studies because of differences in severity between the health states evaluated and the myriad differences in elicitation methods among the four studies.

#### 4.1 Distributional issues

Distributional properties of WTP expressed as absolute dollars are in line with results from other studies. Most studies, along with this one, make note of a positively skewed distribution of WTP expressed in absolute dollars and use non-parametric approaches or mathematical transformations prior to analyses to reduce undue influence of high values. Our skewness statistics, ranging from 2.0–2.9, for monthly WTP expressed in absolute dollars is comparable with skewness statistics from another study in which WTP was elicited using an open-ended format in an interview where participants were asked for their WTP in terms of a “weekly, fortnightly, monthly or yearly figure.” A specific timeframe was not indicated. Skew statistics in that study ranged from 1.7–3.0 (Smith & Richardson, 2005). Even a highly skewed measure is not necessarily invalid, but skewed measures require transformations or use of non-parametric analy-

ses. High values may also indicate that people are giving extraordinarily high values that represent the importance of perfect health without regard for whether they can make the tradeoffs necessary to afford the treatment.

#### 4.2 WTP correlation with income

WTP expressed in absolute dollars clearly has a stronger association with income than WTP expressed in terms of percentage of financial resources. When WTP is expressed as a percentage, the association is negligible for both health states with both percentage formats (this is a natural consequence if participants include their income in considering their financial resources). WTP expressed as absolute dollars showed moderate associations with income. In a recent study, WTP was less sensitive to differences in health state, the higher the proportion of income represented by their WTP because of personal budget constraints (Smith & Richardson, 2005). The extraordinarily high proportion of people giving the same value for both health states when expressing WTP in a single lump sum dollar amount may indicate that a budget ceiling comes into play more readily than with the other 3 formats; i.e., people give a WTP to cure BKA at the maximum of what they can afford and they have no discretionary wealth remaining to cure paraplegia even though they may agree they would be worse off. On the other hand, there is evidence that people are often scale insensitive when giving WTP values — these values may simply reflect the respondent’s subjective desire to be healthy without considering difference in severity (Baron & Greene, 1996).

We have shown that WTP, elicited as a percentage, has superior measurement properties. However, some may argue that we failed to measure what needs measuring (the amount people are willing to pay for various treatment options) with this approach — after all, CBAs require dollars, not percentages. We argue, however, that WTP measured as a percentage can be readily converted to dollar amounts in several ways, and thus provides more flexibility in addition to better measurement properties. As with our study, Schiffner et al. (2003) and Thompson et al. (1984; also Thompson, 1986) found no association between income and WTP when WTP was expressed as a percentage of wealth but, as with many prior studies, we did find that WTP elicited using absolute dollars was moderately and significantly associated with income. The dissociation of WTP from income may be cause for alarm for some economists who regard the presence of this association as one criterion by which to validate the WTP values elicited (Brach et al., 2005; Donaldson, 1999; Donaldson et al., 1997). This may be good news to others, however, who point out the ethical issues that arise when WTP is associated with income — out of

fear that the “buying power” of the rich will give them a disproportionate voice in prioritization schemes (Olsen & Smith, 2001). Some researchers see merit in both concerns (Donaldson et al., 2002).

Percentages can be converted to dollars in two ways. First, for those concerned about the lack of association of income with WTP, percentages can be converted to dollars using individual income (Klose, 1999). Measurement issues aside, these dollars are the same as if elicited directly and thus association with income will be established while preserving the psychometric properties of elicited percentages. In fact, backing into dollars this way may result in WTPs that are more highly correlated with level of income than dollars elicited directly. People may be under-sensitive to their own ability to pay because of the difficulty of thinking about a dollar amount to pay for the good in question and then to consider whether they can afford that amount. The percentage format allows people to think directly in terms of proportion of what they can afford, thus simplifying the task.

Second, those concerned about association of WTP with income have the option of applying the average WTP percentage to average income of the appropriate population (or subgroup) to obtain average WTP in dollars, dissociated with income (Thompson et al., 1984), an approach the World Bank has used to incorporate equity considerations in CBAs of healthcare projects. This approach incorporates distribution weighting consistent with an inequality-averse society (Brent, 2003) for healthcare. Using raw WTP expressed as a proportion of financial resources will result in a group with one-quarter average income having a weight of four while those in an income group with four times the average would have a weight of one-quarter. However, some argue that this approach, at best, results in an “index of the strength of ‘social preferences’” with obscure meaning that makes WTP elicited as a percentage of income irrelevant from the perspective of economic theories underlying the conduct of CBAs (Smith & Richardson, 2005), page 82). Resolving these differing viewpoints and challenges is beyond the scope of this paper.

### 4.3 Limitations and open questions

This study has several limitations. Our scenarios did not specify a timeframe in which payments would need to be made nor how long the cure would last if payment stopped. Though many studies do not spell out specific time-periods (Smith, 2003), it is important to do so to ensure consistent interpretation of the elicitation and results. We conducted this study over the Internet and had a low initial response rate. However, once people clicked onto the survey, 75% of them completed the survey and 98% of those responses were sufficiently valid to include in our

analyses. We did not intend to generalize actual WTP values obtained in this study but rather sought a diverse sample to participate in an experimental study. We were successful in recruiting a diverse sample with respect to age, race and ethnicity, education, and income group. In addition, these demographic characteristics were balanced across the experimental groups. Thus, we expect that the differences we observed in behavior with the four formats in this study will extend to other similar populations. Our results were also in line with those obtained in two pilot studies we conducted using a paper survey of a smaller convenience sample.

WTP expressed as a percentage of monthly financial resources was lower than WTP expressed as a total percentage. Purely mathematically, the percentages should be the same if the same sources of finances were considered in the two timeframes. However, there are many reasons to believe this may not be the case. People may, in fact, be drawing upon different financial resources on a monthly versus lump-sum basis. It would not be unreasonable for respondents to consider the wider range of assets that may be available to them on a one-time lump sum basis. They may be more willing to use their borrowing power or to dip into savings to cure their health condition with a single payment. The monthly timeframe may be more salient for many people who budget on a monthly basis and this format may focus respondents on cash flow where income may be the primary monthly source of incoming cash. Relatively speaking, smaller amounts may be available for discretionary expenditures month-to-month, after paying for things like housing, utilities, and food. Psychologically, shorter timeframes lead to more concrete thinking and predictions (Trope & Liberman, 2003). Though WTP as a percentage of total financial resources performed well based on distributional criteria, we cannot ignore the fact that half of our respondents were willing to forego half or more of their financial resources to cure paraplegia while, on a monthly basis, the median amount was only 30%.

We did not actually convert WTP percentages into dollars for this study. If we did so, based on our data and assuming gross income as the denominator (the only financial measure we collected in this study), values would be significantly higher than dollars elicited directly (for both monthly and annual amounts). Such a comparison, however, is fraught with issues. Dollar amounts would likely be over-estimated because we would not be able to take taxes into account; most people consider after-tax income, not gross income when considering the dollars they can afford to pay for something. However, if people really did consider more than just their income and if we were not constrained by a yearly timeframe, then the converted dollars would be under-estimates. It is clear that more study is needed to discern what respondents are

considering when giving their WTP in dollars or percentages and more elaborate measures of wealth and income are needed. The Health and Retirement Study is one example where participants are asked for information about many components that comprise their financial resources (Juster & Suzman, 1995).

The WTP values elicited in our study were for curing relatively severe disabilities with idealized treatments. Both of these factors led to relatively large, whole number percentages for most participants. But the percentage format may be difficult to use when placing value on more modest (and realistic) treatments. For example, WTP for mammography screening was as low as \$12 in one study (Yasunaga, Ide, Imamura, & Ohe, 2007); it would be very difficult for people to estimate such small a percentage of annual take home income. However, there is evidence that even when eliciting WTP in terms of dollars, low values may be less reliable than high values (Smith, 2006).

More work is needed to determine the validity of responses elicited through the Internet. Though we were concerned about the potential for a high level of protest or spurious responses, we did not see evidence of this. Another study elicited utilities for four different health conditions (including BKA and paraplegia) from this same panel of Internet users who were recruited in the same way at the same time. The large majority of responses were reasonable and valid. Participants gave responses that were highly differentiated between four different health conditions and 74% of those who gave different utilities for BKA and paraplegia (comprising 62% of respondents) gave rankings that were consistent with the corresponding utilities (Damschroder, Zikmund-Fisher, & Ubel, in press). Most of the “questionable” responses in the present study were a result of respondents giving the same non-zero WTP for both health states. The high rate of equal values is troubling, but this may partly be a function of budget constraint (Smith, 2005). The elicitation format appears to influence the rate of inconsistent responses; evident in the lower rate of people with the dollar formats who did not conform to our ordinal criteria compared to the rate for the percentage formats. Many researchers insist that because of the high cognitive demand of WTP elicitation, in-person interviews are necessary (e.g., Arrow et al., 1993). Our results are not much different from another recent study using face-to-face interviews in a large diverse sample in which 41% of participants gave all zeros or equal non-zero WTP values for 3 treatment programs (J.A. Olsen, Donaldson, Shackley, & EuroWill Group, 2005); a reason for some optimism for reliably eliciting WTP values using a web-based instrument.

Nonetheless, the larger question of whether people have consistent values for health conditions with which they are not familiar has yet to be answered definitively.

Regardless of format, further work is needed to determine the appropriate “dose” of information to help people discover what their true preferences are (Watson & Ryan, 2006) – whether coupled with an opportunity for people to deliberate various considerations (e.g., (Abelson et al., 2003; Damschroder, Ubel, Zikmund-Fisher, Kim, & Johri, 2005; Dolan, Cookson, & Ferguson, 1999), feeding back an interpretation of respondent’s WTP so they can affirm or change their response (Watson & Ryan, 2006), or whether researchers simply need better ways to uncover already existing underlying preferences without being influenced by the method (Sugden, 2005). In addition, many psychological questions remain about what WTP elicited using these kinds of methods actually represents. Common sources of biases have been described earlier but in addition, regardless of format, people tend to give the same WTP for varying levels of goods (scale insensitivity), and WTP value for two units valued separately is often higher than WTP for 2 units valued together (lack of additivity) (Baron, 1997), WTP values are often more reflective of perceived market value or cost to produce and not a reflection of their own personal valuation (Baron & Maxwell, 1996). Results from our study help to illuminate ways to elicit consistent and valid WTP amounts from people over the internet, but do not solve the larger issues around WTP values, which despite challenges, continue to be used in CBAs of healthcare programs.

## References

- Abelson, J., Eyles, J., McLeod, C. B., Collins, P., McMullan, C., & Forest, P. G. (2003). Does deliberation make a difference? Results from a citizens panel study of health goals priority setting. *Health Policy, 66*, 95–106.
- Arrow, K., R, S., Portney, P., Leamer, E., R, R., & H, S. (1993). Report of the NOAA panel on contingent valuation. *Federal Register, 58*, 4601–4614.
- Baron, J. (1997). Biases in the quantitative measurement of values for public decisions. *Psychological Bulletin, 122*, 72–88.
- Baron, J., & Greene, J. (1996). Determinants of insensitivity to quantity in valuation of public goods: Contribution, warm glow, budget constraints, availability, and prominence. *Journal of Experimental Psychology: Applied, 2*, 107–125.
- Baron, J., & Maxwell, N. P. (1996). Cost of public goods affects willingness to pay for them. *Journal of Behavioral Decision Making, 9*, 173–183.
- Brach, M., Gerstner, D., Hillert, A., Schuster, A., Sosnowsky, N., & Stucki, G. (2005). Development and evaluation of an interview instrument for the monetary

- valuation of expected and perceived health effects using rehabilitation interventions as a model. *Physikalische Medizin Rehabilitationsmedizin Kurortmedizin*, 15, 76–82.
- Brent, R. (2003). *Cost-benefit analysis and health care evaluations*. Cheltenham, UK: Edward Elgar.
- Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences* (2nd ed.). Hillsdale: Lawrence Erlbaum Associates.
- Damschroder, L. J., Ubel, P. A., Zikmund-Fisher, B. J., Kim, S. Y., & Johri, M. (2005). *A randomized trial of a web-based deliberation exercise: improving the quality of healthcare allocation preference surveys*. Paper presented at the The 27th Annual Meeting of the Society for Medical Decision Making.
- Damschroder, L. J., Zikmund-Fisher, B. J., & Ubel, P. A. (in press). Considering adaptation in preference elicitation.
- Diener, A., O'Brien, B., & Gafni, A. (1998). Health care contingent valuation studies: a review and classification of the literature. *Health Economics*, 7, 313–326.
- Dolan, P., Cookson, R., & Ferguson, B. (1999). Effect of discussion and deliberation on the public's views of priority setting in health care: focus group study. *BMJ*, 318, 916–919.
- Donaldson, C. (1999). Valuing the benefits of publicly-provided health care: does "ability to pay" preclude the use of "willingness to pay"? *Social Science and Medicine*, 49, 551–563.
- Donaldson, C., Birch, S., & Gafni, A. (2002). The distribution problem in economic evaluation: income and the valuation of costs and consequences of health care programmes. *Health Economics*, 11, 55–70.
- Donaldson, C., Thomas, R., & Torgerson, D. J. (1997). Validity of open-ended and payment scale approaches to eliciting willingness to pay. *Applied Economics*, 29, 79–84.
- Haukoos, J. S., & Lewis, R. J. (2005). Advanced statistics: bootstrapping confidence intervals for statistics with "difficult" distributions. *Academic Emergency Medicine*, 12, 360–365.
- Juster, F., & Suzman, R. (1995). An overview of the Health and Retirement Study. *Journal of Human Resources*, 30, S7–S56.
- Kahneman, D., Ritov, I., & Schkade, D. A. (1999). Economic preferences or attitude expressions?: An analysis of dollar responses to public issues. *Journal of Risk and Uncertainty*, 19, 203–235.
- Klose, T. (1999). The contingent valuation method in health care. *Health Policy*, 47, 97–123.
- O'Brien, B., & Gafni, A. (1996). When do the "dollars" make sense? Toward a conceptual framework for contingent valuation studies in health care. *Medical Decision Making*, 16, 288–299.
- Olsen, J. A., Donaldson, C., Shackley, P., & EuroWill Group. (2005). Implicit versus explicit ranking: On inferring ordinal preferences for health care programmes based on differences in willingness-to-pay. *Journal of Health Economics*, 24, 990–996.
- Olsen, J. A., & Smith, R. D. (2001). Theory versus practice: a review of "willingness-to-pay" in health and health care. *Health Econ*, 10, 39–52.
- Payne, J. W., Bettman, J. R., & Schkade, D. A. (1999). Measuring constructed preferences: Towards a building code. *Journal of Risk and Uncertainty*, 19, 243–270.
- Schiffner, R., Schiffner-Rohe, J., Gerstenhauer, M., Hofstadter, F., Landthaler, M., & Stolz, W. (2003). Willingness to pay and time trade-off: sensitive to changes of quality of life in psoriasis patients? *Br J Dermatol*, 148, 1153–1160.
- Smith, R. D. (2000). The discrete-choice willingness-to-pay question format in health economics: Should we adopt environmental guidelines? *Med Decis Making*, 20, 194–206.
- Smith, R. D. (2003). Construction of the contingent valuation market in health care: a critical assessment. *Health Econ*, 12, 609–628.
- Smith, R. D. (2005). Sensitivity to scale in contingent valuation: The importance of the budget constraint. *Journal of Health Economics*, 24, 515–529.
- Smith, R. D. (2006). The relationship between reliability and size of willingness-to-pay values: a qualitative insight. *Health Economics*, 9999, n/a.
- Smith, R. D., & Richardson, J. (2005). Can we estimate the "social" value of a QALY? Four core issues to resolve. *Health Policy*, 74, 77–84.
- Sugden, R. (2005). Anomalies and Stated Preference Techniques: A Framework for a Discussion of Coping Strategies. *Environmental and Resource Economics*, 32, 1–12.
- Thompson, M. S. (1986). Willingness to pay and accept risks to cure chronic disease. *Am J Public Health*, 76, 392–396.
- Thompson, M. S., Read, J. L., & Liang, M. (1984). Feasibility of willingness-to-pay measurement in chronic arthritis. *Med Decis Making*, 4, 195–215.
- Trope, Y., & Liberman, N. (2003). Temporal construal. *Psychological Review*, 110, 403–421.
- Venkatachalam, L. (2004). The contingent valuation method: a review. *Environmental Impact Assessment Review*, 24, 89–124.
- Watson, V., & Ryan, M. (2006). Exploring preference anomalies in double bounded contingent valuation. *J Health Econ*.
- Whynes, D. K., Frew, E. J., & Wolstenholme, J. L. (2005). Willingness-to-pay and demand curves: A comparison of results obtained using different elicita-

tion formats. *International Journal of Health Care Finance Economics*, 5, 369–386.

Whynes, D. K., Wolstenholme, J. L., & Frew, E. (2004). Evidence of range bias in contingent valuation payment scales. *Health Economics*, 13, 183–190.

Yasunaga, H., Ide, H., Imamura, T., & Ohe, K. (2007). Women's anxieties caused by false positives in mammography screening: a contingent valuation survey. *Breast Cancer Research and Treatment*, 101, 59–64.

Yeung, R. Y., Smith, R. D., Ho, L. M., Johnston, J. M., & Leung, G. M. (2006). Empirical implications of response acquiescence in discrete-choice contingent valuation. *Health Economics*, 15, 1077–1089.

## Appendix: Health state descriptions

### Below-the-knee amputation (BKA)

Imagine that you have a below-the-knee amputation and have gone through the rehabilitation process. You use a prosthetic device, an artificial leg that fits well and is fairly comfortable. Walking requires more effort, but you get around pretty well and have only a slight limp. When you are wearing long pants, nobody can tell that you are using a prosthesis. Because your amputation is below the knee, you can still participate in sports activities; you just won't be able to run as fast or jump as high. Other than your amputation, you are perfectly healthy.

### Paraplegia

Imagine living with paraplegia. Your legs are paralyzed from the waist down. You cannot move your legs and you have to use a wheelchair to get around. Your bladder and bowel functioning are both normal; however, you sometimes need help getting to the toilet. You also require help in bathing and other daily activities. You do not have any health problems other than paraplegia.