2006

A comparison of alternative certainty calibration techniques in contingent valuation

M Samnaliev

TH Stevens

T More

Analysis

A comparison of alternative certainty calibration techniques in contingent valuation

Mihail Samnaliev [a,*], Thomas H. Stevens [b], Thomas More [c]

[a]Center for Health Policy and Research, University of Massachusetts Medical School, Chang Building, 222 Maple Ave., Shrewsbury, MA 01545, United States
[b]Department of Resource Economics, University of Massachusetts, 80 Campus Center Way, Amherst, MA 01003, United States
[c]USDA, Forest Service, Northeastern Research Station, P.O. Box 968, Burlington, VT 05402, United States

Received 25 August 2003; received in revised form 2 May 2005; accepted 2 May 2005
Available online 12 September 2005

Abstract

A field test of two types of certainty calibration techniques in contingent valuation of public lands indicated that a 10-point certainty scale reduced WTP estimates by about half. Adjusting for uncertainty via a 'Not Sure' option did not reduce WTP estimates but the variance increased. There are several differences between these two ways of accounting for respondents' uncertainty, which may suggest why they provide different WTP value estimates and variances.

Keywords: Certainty calibration; Contingent valuation; CVM; WTP

1. Introduction

Empirical evidence suggests that significant uncertainty often exists in responses to contingent valuation questions (Alberini et al., 2003). Since respondent uncertainty has often been related to the problem of hypothetical bias (see Harrison and Rustrom, in press; List and Gallet, 2001), several contingent valuation, CVM, formats that allow respondents to express uncertainty directly have been developed. Examples include the multiple-bounded question format (Welsh and Poe, 1998), a "random-valuation" model (Wang, 1997), various uncertainty scales (Champ et al., 1997; Ekstrand and Loomis, 1997) a polychotomous choice format (Ready and Navrud, 1999), and NOAA's well-known 'Don't Know' or 'Not Sure' option. However, agreement about the appropriate method for uncertainty adjustment is far from universal. For example, Wang (1997), Carson et al. (1994), and Alberini et al. (2003) present very different views about calibration for uncertainty.

A 10-point certainty scale following a dichotomous choice, DC, format and the inclusion of a 'Not Sure'

* Corresponding author.
E-mail addresses: Mihail.Samnaliev@umassmed.edu (M. Samnaliev), tstevens@resecon.umass.edu (T.H. Stevens).

doi:10.1016/j.ecolecon.2005.05.017

option are two common ways to incorporate uncertainty. Use of a certainty scale with a cut-off point of 8 and 10 (with 10 being very certain) has been shown to provide similar hypothetical and actual willingness to pay, WTP, estimates (Champ et al., 1997). The treatment of 'Not Sure' responses has been more controversial (Wang, 1997), but a common approach has been to treat them as either 'No' or missing (Alberini et al., 2003; Carson et al., 1994).

This study compares the effect of these two types of certainty adjustment on WTP estimates in a randomized split sample mail survey. We find that treatment of 'Yes' responses with certainty of less than 8 (or 10) as 'No' provide different willingness to pay estimates than treatment of 'Not Sure' responses as either 'No' or as missing. We then contribute to the discussion on the motivation underlying uncertain responses and argue that the two calibration methods may be conceptually different.

2. Previous studies

The motivation behind uncertain responses is not well understood. After the NOAA panel suggested that a 'Don't Know' option should be added to the DC CVM format, a body of literature has explored respondent motivation underlying 'Not Sure' responses. Alberini et al. (2003) suggest three interpretations of responses to this option. One possibility is that 'Don't Know' respondents are not in the market for the good being valued. A second interpretation is that 'Don't Know' respondents have not yet made up their mind. The third possibility is that these responses reflect uncertainty. Moreover, Alberini et al. define two types of uncertainty: (a) "true" uncertainty wherein respondents have insufficient experience and (b) "false" uncertainty wherein respondents do not want to spend time thinking about the valuation question or would like to indicate some support for the item being valued, but would not pay the amount asked. Carson et al. (1994) recommend that 'Not Sure' responses be treated as missing, because respondents who choose the 'Not Sure' option would say 'No' if actually forced to choose. In addition, Champ et al. (2003) find that respondents may choose the 'Not Sure' option because they are uncertain about their income, ability to commit to spending money, or about the benefits of the program. Other hypotheses include the notion that uncertainty may arise because of lack of knowledge, interest, or inability to make a quick decision.

Wang (1997) presented an alternative interpretation of 'Don't Know' responses. He argued that 'Don't Know' (or 'Not Sure') answers represent the point of indifference to the offered bid. As the price of the commodity increases, a typical respondent would switch her answer from 'Yes' to 'Don't Know' and from 'Don't Know' to 'No'. Wang included the 'Don't Know' answers in a multinomial probit model estimation and concluded that they provide useful information about preferences.

On the other hand, certainty scale calibration has become quite popular in dichotomous choice (DC) CV studies. In this approach, people are asked how certain they are of their response on a 10-point scale. A common application of the certainty scale is to treat positive answers as 'Yes' only when certainty levels are at least 8 on a 10-point scale with 10 indicating 'Very Certain' (for example, see Champ et al., 1997). The effectiveness of this method has been established by comparing hypothetical payments to actual donations (Champ et al., 1997; Polasky et al., 1996). These, as well as other recent studies, suggest that uncertainty scale calibration can reduce hypothetical bias and/or so called 'Yea-Saying' effects. However, Ekstrand and Loomis (1997) reported that the effect of this method depends on how the scale is used. Bias reduction was reported when certainty levels of at least 8 were used to calibrate only 'Yes' answers, but reduction of bias was questionable when 'No' answers were also calibrated. In addition, the authors found that certainty calibration reduced the goodness of fit (of the logit WTP model) and increased the variance in responses.[1]

Taken together, these arguments demonstrate the complexity of the issue of uncertainty calibration. Uncertainty is not a precise or single condition and may be caused by a range of factors. Further, little is known about the separate or confounding effect of each factor and this presents a methodological prob-

[1] However, Welsh and Bishop (1993) reported that certainty calibration reduced the variance in responses. Several other studies have also applied certainty scales to calibrate 'Yes' and 'No' responses (Li and Mattsson, 1995).

lem in CVM applications. Inclusion a of 'Not Sure' option or use of a DC format followed by a certainty scale have been the most common methods to empirically account for uncertainty. Given that these two formats are different, a natural question is whether they produce the same WTP estimates.

Our study employs a split sample where each group has similar socio-economic characteristics and are presented with identical hypothetical settings. This allows testing for differences between a certainty scale and a 'Not Sure' option regardless of underlying motivations related to uncertainty, which are assumed to be, on the average, identical between the two samples.

3. Methods

A mail survey was used to elicit attitudes towards user fees to access public lands in the context of the current US Fee Demonstration Program (FDP). The FDP has been experimentally implemented for some public lands and allows several US agencies, including the Forest Service, the Bureau of Land Management, and the US Fish and Wildlife Service, to impose access fees for public use of these lands. The purpose of the FDP is to test the appropriateness of entrance fees as a mechanism to raise additional money to maintain public natural resources and recreation sites.

The survey was pre-tested with a pilot survey in June 2002 and then mailed in October to about 1600 randomly selected households in New Hampshire and Idaho. Within each state, a two-stage cluster sampling was applied in order to distinguish between the urban and rural population. In an effort to increase response rates and reduce non-response bias, we followed the four-step procedure proposed by Dillman (2001). The overall response rate was 34%, for a total of 540 observations.

The hypothetical valuation part of the survey consisted of a description of a recreation area (a hypothetical public site with a scenic overview), which had become part of the FDP. The willingness to pay (WTP) question was presented in a DC format in which respondents were asked to make hypothetical payments of randomly assigned prices ($3, $5, or $10) for access to this site. Two versions of the questionnaire were mailed. The first was a baseline version consisting of 'Yes' and 'No' options followed by a standard 10-point certainty scale. The second version (version NS) included a 'Not Sure' category for the WTP response (see Appendix A). Both versions asked for hypothetical payments.

The theoretical utility model and the derivation of willingness to pay follow well-established procedures, outlined in Appendix B. Mean WTP was calculated by integrating under a logit function where price was truncated at $25 and bounded to be positive:

$$\text{mean WTP} = \int_0^{25} \left[1 - G_{\text{wtp}}\right] \mathrm{d}W. \tag{1}$$

We first present the unadjusted distribution of WTP across survey versions and then use logit models to control for effects of a set of associated variables and to calculate mean WTP and 95% confidence intervals.

4. Results

Variables included in the analysis are described by survey version in Table 1. Two-sample t-tests for difference in means and proportions showed that the distribution of variables and respondent characteristics between survey versions were statistically indistinguishable, as expected, since survey versions were mailed randomly. This allows evaluation of the effects on WTP that arise due to different treatments and eliminates the possibility of confounding effects due to differences between "average" respondents to each survey version.

Respondents had an average income of between $45 and $60 thousand per year per household and most had at least a college degree.[2] The size of most households varied between 1 and 4, with an average of 2.7 per household. Average age was 56.5 years, skewed towards the upper tail of the population distribution. About 58% of the respondents reported visiting public lands at least three times a year over

[2] This suggests that our sample may not be representative of the general population, but we do not expect this to affect the results in a methodological study which compares two 'treatment' effects.

Table 1
Descriptive characteristics ($n=540$)

Variable	Description	Coding	All versions	Mean (S.D.)	
				Baseline version $n=281$	NS version $n=259$
Price ($)	Dollar amount asked in the questionnaire	$3, $5, $10	5.9 (2.9)	6.05 (2.9)	5.86 (2.9)
Income	Yearly income, coded in 10 categories	1 ($10,000) to 10 (>$120,000)	5.2 (2.2)	5.3 (2.3)	5.1 (2.1)
Visits	Whether respondents visited public lands more than three times a year in the past 3 years	1=Yes 0=Otherwise	0.58 (0.49)	0.56 (0.49)	0.60 (0.49)
HH	Number of household members	Continuous	2.7 (1.6)	2.64 (1.3)	2.82 (2.0)
State	Whether resident of NH or Idaho	0=NH 1=Idaho	0.51 (0.3)	0.52 (0.2)	0.50 (0.5)
Age	Age of respondent	Continuous	56.5 (15.01)	57.2 (14.9)	55.7 (14.7)
Round	Whether survey was returned in first round or in second round	0=first round 1=second round	0.27 (0.42)	0.27 (0.40)	0.28 (0.45)
Urban	Whether survey was sent to an urban or a rural cluster	0=rural 1=urban	0.09 (0.25)	0.08 (0.2)	0.10 (0.3)

the last 3 years.[3] The mean number of visits in the past 3 years was 11 visits per person.

The unadjusted distribution of 'Yes', 'No' and 'Not Sure' responses for each survey version and by price level is shown in Table 2. About 62% rejected the fee offered in the baseline version and about half of the respondents rejected the fee in the NS version. As expected, the proportion of respondents who were willing to pay the proposed fee decreased as price increased, and as the certainty scale was applied to the baseline version. Approximately one in six chose the 'Not Sure' option in the NS version.[4]

Three groups of variables were included in the willingness to pay model, based on theoretical expectations from classical economics and regardless of their statistical significance: (i) dollar amount requested and income, (ii) individual tastes and preferences, and (iii) social characteristics, respectively. These were represented by the variables price, (previous) visits to recreation lands, age, and household size.

We hypothesize that residents of Idaho and New Hampshire differ culturally in their preferences, and that residents of rural and urban areas differ in their lifestyle regarding outdoor activities. The effect of these two factors were represented by the variables 'state' and 'urban'. We also included a variable accounting for the round (time frame) in which the surveys were returned. Linearity of age and income was examined visually by plotting these variables on a logit scale. The inclusion of the variables state, round, and urban was then assessed on the basis of three

Table 2
Distribution of willingness to pay

Version	All	$3	$5	$10
Baseline	281	97	95	89
WTP=No	175 (62.3%)	48 (49.5%)	57 (60.0%)	70 (78.7%)
WTP=Yes	106 (37.7%)	49 (50.5%)	38 (40.0%)	19 (21.3%)
Certainty8				
WTP=No	216 (76.9%)	66 (68.0%)	71 (74.7%)	79 (88.8%)
WTP=Yes	65 (23.1%)	31 (32.0%)	24 (25.3%)	10 (11.2%)
Certainty10				
WTP=No	239 (85.1%)	79 (81.4%)	78 (82.1%)	82 (92.1%)
WTP=Yes	42 (14.9%)	18 (18.6%)	17 (17.9%)	7 (7.9%)
Not Sure (NS)	259	86	83	90
WTP=No	129 (49.8%)	35 (40.7%)	38 (45.8%)	56 (62.2%)
WTP=Yes	83 (32.1%)	32 (37.2%)	30 (36.1%)	21 (23.3%)
WTP=Not Sure	47 (18.1%)	19 (22.1%)	15 (18.1%)	13 (14.5%)

[3] This criteria was chosen arbitrarily. We assume that people who visit public lands at least three times a year are regular visitors who have well-formed preferences for public lands, while respondents who visit public lands occasionally, say once a year, may not have well-established preferences.

[4] One possible reason for this rather high level of 'Not Sure' responses is that we provided relatively little detail about the commodity being valued in this study (see Appendix A). Although lack of detail may mean that many respondents interpreted the payment question as about paying user fees in general as opposed to payment for a specific site, it is one way to introduce "demand" uncertainty.

criteria: (1) significance in a univariate model as a main effect variable, (2) likelihood ratio test after inclusion in the main effects model, and (3) the effect of the variable as a modifier on the other variables (percentage change of the estimated coefficients). Interaction terms were considered on the basis of plausibility and statistical significance. Several versions of a logit model were specified wherein willingness to pay (Yes = 1, No = 0) was regressed on the variables listed in Table 1. A likelihood ratio test for difference of estimated coefficients between the two states showed that the estimates were not statistically different which allowed the data to be pooled.

4.1. Estimation of WTP

Logit models were estimated for each survey version. Estimates for the baseline and the NS treatment (where 'Not Sure' responses were treated as missing) are shown in Table 3. Results from the logit models were consistent with the unadjusted results in Table 2 with one exception, the proportion of 'Yes' responses in the Baseline version (37.7%) was higher than in the NS version when Not Sure responses were treated as 'No' (32.1%), but mean WTP was greater in the latter version. This can be attributed to the relative distribution of the responses by price and to the slightly uneven effect of the explanatory variables across versions, which were unaccounted for in Table 2.

Table 3
Logistic estimation of WTP function

Variable (expected sign)	Baseline version	NS version (treating not sure as missing)
	Estimate (standard error)	Estimate (standard error)
Intercept	3.30 (2.00)[a]	2.34 (2.47)
Price (–)	– 0.26 (0.05)[b]	– 0.17 (0.05)[b]
Income (+)	0.18 (0.08)[c]	0.15 (0.08)[a]
Visits (?)	– 0.72 (0.30)[c]	– 1.20 (0.37)[b]
HH (–)	– 0.23 (0.11)[c]	–0.23 (0.16)
State (?)	0.20 (0.29)	0.16 (0.33)
Age (?)	– 0.13 (0.06)[c]	– 0.06 (0.07)
Age^2	0.001 (0.0005)[c]	0.0005 (0.0006)
Round (?)	0.45 (0.32)	0.40 (0.37)
Urban (?)	0.23 (0.46)	0.30 (0.58)

[a] Significant at 90% level.
[b] Significant at 99% level.
[c] Significant at 95% level.

Table 4
Mean willingness to pay and 95% confidence intervals (1000 bootstraps)

Version	Mean WTP ($) 95% CI	Difference between upper and lower CI	Baseline mean WTP/ Version mean WTP
Baseline (DC format) $n = 281$	4.32 3.32–5.46	2.14	1.0
Certainty8 (WTP = Yes only for certainty ≥ 8) $n = 259$	2.65 1.72–3.59	1.87	1.63
Certainty10 (WTP = Yes only for certainty = 10) $n = 259$	1.68 0.79–2.67	1.88	2.57
NS 'Not Sure' = missing $n = 182$	5.43 3.57–9.11	5.54	.80
'Not Sure' = Yes $n = 224$	7.28 5.25–11.02	5.77	.59
'Not Sure' = No $n = 224$	4.87 2.69–10.58	7.89	.89

The effects of price, income, and number of household members (see Table 3) were as expected: positive effect of income and negative effect of permit price and number of household members. We did not have prior expectations for the effect of the variable 'visits'. Visitors of public lands might be expected to be more likely to pay, since they are the users of the commodity that is being valued. However, in this particular study, users may be less likely to pay because of strategic objection to user fees. Our estimates show a negative effect for this variable. The effect of round can be positive or negative. People who are less interested in public lands can be expected to respond later, which means that the effect of round would follow the same logic as the effect of previous visits. However, we might expect later respondents to be mainly working people with busier schedules, which might suggest a positive effect of the variable round.

Mean WTP values were calculated using Eq. (1). Previous research has suggested two main methods for confidence interval estimation. Park et al. (1991) proposed a simulation method based on the Krinsky and Robb (1986) technique where a Gauss distribu-

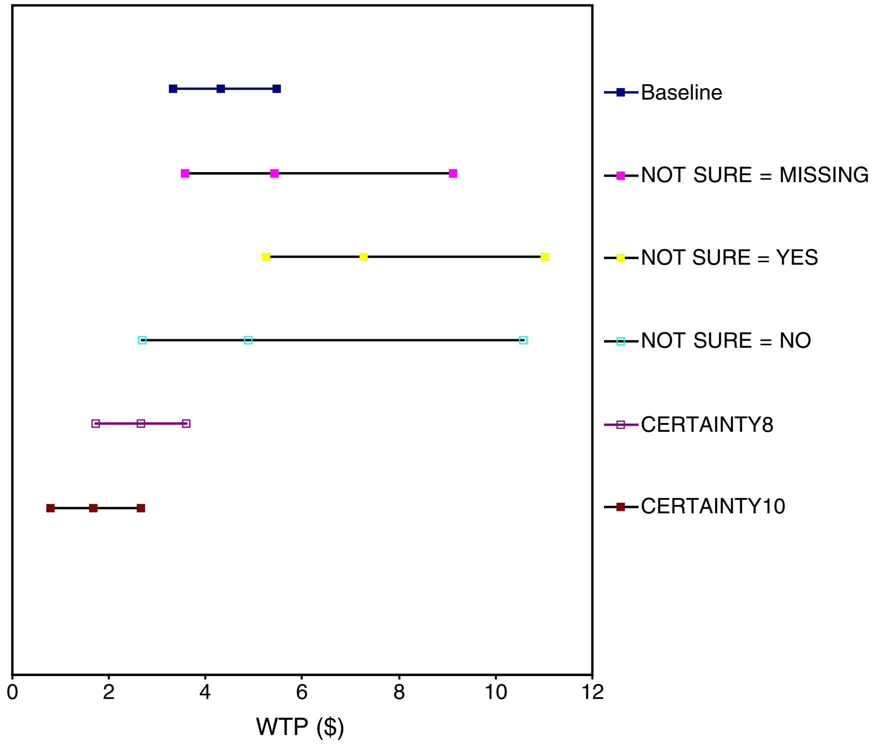

Fig. 1. Mean WTP and 95% confidence intervals for each version WTP ($).

tion is simulated around each estimated coefficient using its estimate and variance. The second approach, proposed by Duffeld and Patterson (1991), is based on bootstrapping (with replacement) from the original sample. We use the second approach, which, as pointed out by Cooper (1994), does not impose normality on the distribution of the coefficients. Bootstrapping was done in SAS. Mean WTP was calculated through integration using MATHEMATICA. Empirical confidence intervals around the point estimate of mean WTP were constructed by generating 1000 bootstraps with replacement for each version.

In order to test for the relative effect of each type of uncertainty calibration, we compared the estimates of mean WTP obtained from the baseline version to the estimates of mean WTP derived from the version NS, Certainty8, and Certainty10.[5] Mean WTP and confidence intervals for all versions are presented in Table 4 and plotted in Fig. 1. In addition, we compared the probabilities of a 'Yes' response for each fee level in order to test whether certainty calibration might have a different effect as fee asked for increases (see Table 5).

4.2. *Effect of certainty scale and 'Not Sure' option*

The effect of certainty scale and 'Not Sure' calibration was tested by comparing mean WTP estimates (Table 4, column 2 and Fig. 1) and confidence intervals (Table 4, column 2 and Fig. 1) to the baseline. The certainty scale versions produced a lower mean WTP relative to the baseline by a factor 1.6 in the Certainty8 version and by a factor of 2.6 in the Certainty10 version. These results are to be expected since in both cases some 'Yes' responses are recoded as 'No'. The mean WTP confidence interval associated with the Certainty8 version overlaps the baseline version, but mean WTP confidence intervals for the Certainty10 and baseline version do

[5] In the Certainty8 and Certainty10 versions, all 'Yes' responses in the baseline version followed by certainty of less than 8 or 10, respectively, were recoded as 'No'. In the NS version, 'Not Sure' responses were treated in three ways; 'No', 'Yes', or 'Missing'.

Table 5
Distribution of certainty levels, n (%)

Certainty	All $n=260$	\$3 $n=90$	\$5 $n=90$	\$10 $n=80$	Yes $n=108$	No $n=160$
<8	63 (24%)	23 (25%)	21 (24%)	19 (23%)	33 (35%)	29 (18%)
8–9	62 (24%)	23 (26%)	20 (22%)	19 (24%)	23 (23%)	39 (25%)
10	135 (52%)	44 (49%)	49 (54%)	42 (53%)	42 (42%)	92 (57%)

not overlap. On the other hand, mean WTP was higher relative to the baseline for all NS treatments (see Table 4). Importantly, our data suggest that calibration of certainty through the traditional recoding of 'Yes' responses with certainty of 8 (mean WTP=\$2.65) or 10 (mean WTP=\$1.68) as 'No' provides different results as compared to a 'Not Sure' calibration when 'Not Sure' responses are treated as 'No' (mean WTP=\$4.87) or missing (mean WTP= 5.43). In addition, the variation of WTP values was much smaller for the first two estimates (see Table 4 and Fig. 1).

It is also worth noting that certainty scale can be applied in a variety of combinations: calibrating only Yes or only No responses with certainty <8 or <10 and treating them as No (or Yes, respectively) or as missing, or calibrating both Yes and No responses (which yields 10 different ways to adjust WTP estimates through a certainty scale). Any of these calibrations would naturally produce different estimates, as it is applied to the same sample of respondents. However, previous findings (for example, Champ, et al., 1997) suggest that hypothetical payments are similar to actual payments only when 'Yes' responses are calibrated.

5. Why do they differ?

The different WTP results derived from the two ways of adjusting for uncertainty may be attributed to several factors. When viewed from a simple empirical perspective, the usual uncertainty adjustment employed in this and in most other CVM studies treats uncertain responses as "no". In the case presented here, 48% of respondents were at least somewhat uncertain (certainty less than 10) and 24% gave a certainty level of less than 8. However, only 18.1% selected "not sure". This means that a larger proportion of individuals are treated as giving a "no" response in the uncertainty adjustment method as compared to the NS format. Consequently, mean WTP estimates derived from the NS method are higher than those obtained from the uncertainty adjusted method.[6]

Moreover, as can be seen in Table 2, by looking at the proportions of responses between the Baseline and the Not Sure (NS) samples, as price increases, more people who responded 'Not Sure' in the NS version, would have chosen 'No' if they were given the option of only saying Yes/No. That is, at a \$10 bid of those giving the "not sure" option, 14.5% chose "not sure" and 62.2% said "no" which when added together equals 76.7% which is pretty close to the 78.7% of 'Nos' in the baseline version. Similarly, at \$5, the sum of 'Not Sure' (18.1%) and 'No' (45.8%) is close to the "no" proportion in the Baseline (60.0%). However, at a \$3 bid, 'Not Sure' responses would fall more evenly between 'Yes' and 'No'. Further, the percent of respondents who select the 'Not Sure' option declines as price increases. Thus, at large bids 'Not Sure' seems to capture 'No' responses, while at small bids, it seems to capture both 'No' as well as 'Yes' responses. This result may suggest that 'Not Sure' responses need to be treated differently as price increases and converting 'Not Sure' responses to 'No' may be justified at high bid levels (\$5 and \$10 in this study) but may not be justified at low bid levels (\$3 in this study), where it would make more sense to distribute them between both 'Yes' and 'No' responses.

[6] Also, as noted by one reviewer, suppose that 50% of respondents selecting the NS or uncertain (≥ 8) categories are really "no" and 50% are really "yes". By treating all uncertain respondents as "no" (which is traditional), estimated WTP derived from the NS format when "not sure" responses are treated as missing will generally be greater than the WTP derived from the uncertainty adjusted format.

A more theoretically based explanation for why the uncertainty adjustment and NS formats used here might produce different results is the possibility of "yea-saying" by respondents. As noted by Brown et al. (1996), Blamey et al. (1999), and Holmes and Kramer (1995), yea-saying appears to play a significant role in many CVM studies. Kanninen (1995), for example, estimated that 20% of her respondents were yea-sayers, and many researchers have suggested that yea-saying may be a factor associated with hypothetical bias.

Yea-saying is generally assumed to be linked with uncertainty (see Champ et al., 1997) and in our study about 58% of respondents who were willing to pay something were also uncertain (see Table 5). These respondents probably did not have an exact estimate of their WTP, and as a result, they may have anchored on the posited bid amount. When given a dichotomous choice of either "yes" or "no", yea-saying may therefore result in a greater proportion of yes responses to each bid amount. This behavior may be particularly relevant in hypothetical situations where yea-saying is essentially costless.

On the other hand, when given a format wherein uncertain respondents have a choice of "yes", "no", or "not sure", some so-called yea-sayers may tend to select the not sure option while others may still anchor on the given bid amount and respond "yes". Consequently, when a not sure option is available, there may be fewer "no" responses than otherwise and estimated willingness to pay will be higher than the baseline.[7] Since 49.8% of our respondents said no in the NS format while 62.3% said no in the baseline, estimated mean WTP derived from the NS format was generally greater than that associated with the baseline (see Tables 2 and 4).

Another factor is that some respondents might have been influenced by the way the survey was worded. For example, "no" and "not sure" in the NS version were followed by a request to "please explain", but the "yes" choice was not (see Appendix A). Because of this, the "no" and "not sure" response may have appeared to go together, and this may have increased the likelihood that people who were inclined to choose "no" would have chosen "not sure" relative to those who said "yes".

Another consideration is that the underlying motivation for a 'Not Sure' choice may differ from the motivation for choosing a low level of certainty on a 10-point scale. If so, then applying these two approaches to identical samples might produce different WTP estimates because these capture different types of uncertainty.

While there is a fair amount of literature on the motivation behind 'Not Sure' responses, the motivation behind uncertain responses when a scale is used has not been widely discussed. Some possible hypotheses are outlined below. The implicit assumptions when certainty scales are being used to calibrate 'Yes' responses (for example, as in Champ et al., 1997) can be summarized in Hypothesis I. Hypothesis II is based on Wang's argument. Two other factors that may play an important role in uncertainty adjustment of a DC CV question are suggested in Hypotheses III and IV.

Hypothesis I. Self-reported certainty to a 'Yes' response provides information about the individual's true utility-maximizing price. A respondent who overstates his WTP (due to 'Yea-Saying'; for example, in a DC format) calibrates his response, using the certainty scale, until he reaches the optimal price. Certainty to 'No' responses does not yield any relevant information about one's WTP.

Hypothesis II. Certainty is lowest at the price that is the true willingness to pay (Wang's argument). In this study, since the mean WTP is about \$5, we can expect that at \$5, the average certainty level would be smaller as compared to \$3 and \$10.

Hypothesis III. Certainty represents consistency between answers. People tend to avoid personal contradictions, and once they choose a 'Yes' or a 'No' response, they tend to back it with a high level of certainty.

Hypothesis IV. Certainty represents a general attitude about the program being valued, rather than economic value. By indicating high levels of certainty to a 'Yes' response, respondents may be expressing their support of the program being valued. By marking high levels

[7] In this context, it is interesting that the proportion of not sure respondents declines as the bid amount increases (see Table 2).

Table 6
Distribution of certainty levels by gender and attitudes about the FDP

Certainty	Males (%)	Females (%)	Objected to fees on principal (%)	Did not object to fees (%)
8–10	78.8 (p=0.06)[a]	67.6	83.3 (p=0.09)[a]	73.2
10	53.8 (p=0.2) [a]	45.1	71.2 (p=0.0003)[a]	45.4

For example, Ho: 78.8= 67.6.
[a] Kruskal–Wallis test.

of certainty to a 'No' response, they may express objection in principle.

Our data did not seem to support Wang's hypothesis. The proportion of respondents who were certain was not lowest at the \$5 price. Hypothesis III is based on the theory of stability and is related to the notion that people avoid cognitive dissonance in their responses (for example, see Schwarz and Sudman, 1996). Even if a person hesitates about whether to say 'Yes' or 'No', when asked later about her certainty, she would tend to indicate a high level of certainty to avoid self-contradiction. In this case, unlike 'Not Sure' certainty scale responses would enhance the 'Yes' or 'No' with increasing magnitude along the scale. Consistency may be expressed during in-person interviews or be an internalized norm of behavior that appears regardless of social settings.

Given this premise, we hypothesize that certainty levels may represent consistency between answers rather than true WTP or true 'Not sure'. Although a sound test of this hypothesis is beyond the scope of this article, an intuitive consequence would be that the distribution of certainty levels would be skewed towards 10. Our data did show an uneven distribution along the certainty scale with certainty levels strongly skewed towards 10 and very few certainty levels less than 5. About half of the responses were followed by a certainty level of 10 indicating 'Very Sure'. This was the case for all price levels and for both 'Yes' and 'No' responses (see Table 5).

Hypothesis IV argues that certainty of a WTP response may be a manifestation of attitude, rather than true willingness to pay. In order to test this hypothesis, we explored the association between certainty levels and attitudes towards user fees. Respondents who objected to fees in principle were more certain in rejecting the price asked. Males tended to be more certain in their answers than women, but this result is mixed; the significance of gender depended on how certainty was coded. The associations of certainty levels to gender and attitudes about user fees are summarized in Table 6. Certainty to 'No' responses was greater among those who objected to fees, implying that high certainty to a 'No' response is a way to assert objection. Certainty to 'Yes' responses was not correlated with price or attitudes.

Table 7 presents logistic regression estimates in which certainty is regressed on objection to fees, gender, and whether price was \$5 (which is the average WTP) or not. Among 'No' responses, negative attitudes towards user fees had a highly significant effect on certainty levels. Respondents who objected to user fees in principle were on average

Table 7
Predictors of certainty levels

Variable (expected sign)	Certainty of 10 to a 'No' response (n=157)		Certainty of 10 to a 'Yes' response (n=97)	
	Estimate (standard error)	Odds ratio (95% CI)	Estimate (standard error)	Odds ratio
Intercept	0.35 (0.54)		0.35 (0.54)	
Mean price (−): 1 if price=\$5; 0 otherwise	0.15 (0.35)	1.2 (0.6–2.3)	0.30 (0.43)	1.4 (0.6–3.2)
Object (+): 1 if objected to user fees in principal; 0 otherwise	1.01[a] (0.35)	2.7[a] (1.4–5.5)	0.5 (1.4)	1.6 (0.1–27.7)
Gender (+): 1=Male; 2=Female	−0.38 (0.37)	0.7 (0.3–1.4) LR=0.01	−0.03 (0.37)	1.0 (0.4–2.6) LR=0.9

[a] Significant at 99% level.

2.7 times more likely to indicate certainty of 10. The effect of gender was insignificant as was the effect of price.

6. Conclusion

In a mail contingent valuation survey utilizing a randomized split sample the two common ways of calibrating for uncertainty, a certainty scale where ‘Yes’ responses are recoded as ‘No’ and a ‘Not Sure’ option recoded as ‘No’ or missing, generally produced different results. While it is challenging to attribute this difference to a single factor, data analysis pointed to several possible explanations, including the presence of a ‘Yea-Saying’ effect and conceptual difference between ‘Not Sure’ and a certainly scale. At high bid levels ‘Not Sure’ responses seem to represent ‘No’ responses, while at low bid levels ‘Not Sure’ responses represented both ‘Yes’ and ‘No’. This would suggest that converting ‘Not Sure’ responses to ‘No’ might only be justified at high bid levels ($5 and $10 in this study) but not at low bid levels ($3 in this study). Further, when a certainty scale is used, high levels of certainty may be an indication of consistency between answers where people reinforce their ‘Yes’ or ‘No’ responses. High levels of certainty may also represent expressions of attitudes rather than monetary values. In the latter case, certainty scales may not be able to consistently reduce hypothetical bias across applications of CVM.

Further research on the relationship between certainty scale levels and individual characteristics is needed in order to verify the validity of this technique and its ability to consistently reduce or eliminate hypothetical bias. Finally, our analysis was based on hypothetical payments elicited in a mail survey and future research incorporating real payments to compare the two ways of certainty calibration may be informative.

7. Uncited reference

National Oceanographic and Atmospheric Administration (NOAA), 1994

Acknowledgements

This research was supported in part by funds provided by the USDA, Forest Service, Northeastern Research Station.

Appendix A

Hypothetical Settings

Imagine an area with a scenic overlook in a nearby federal or state public forest. In the past, this area was free with only picnic tables and a dirt parking lot. This year the area is the same as always, but it is part of the Fee Demonstration Program (described in the cover letter), so you must buy a permit or face a fine of $100 if caught without a permit. Permits are sold at a visitor's center that you pass on the way to the site.

If a permit to use this area costs $______ per visitor per day, would you buy it, keeping in mind your household income and other financial commitments?

Baseline Version

A. Yes, I would pay this amount.

B. No, I would not pay this amount. (Please explain why)

Certainty Calibration

C. How sure are you of your decision about how much you would pay? Please circle one number from 1 to 10, with 1 indicating "very unsure" and 10 indicating "very sure".

1	2	3	4	5	6	7	8	9	10
very unsure									very sure

If you are unsure, please explain why.

Version Certainty8: All 'Yes' responses followed by certainty less than 8 are recoded as 'No'

Version Certainty10: All 'Yes' responses followed by certainty less than 10 are recoded as 'No'

Version NS

A. Yes, I would pay this amount.

B. No, I would not pay this amount. (Please explain why)

C. Not Sure. (Please explain why)

Appendix B

We follow a simple model where individual utility is a function of Income, *Y*, a basket of market goods, *X*, and public land use, *Q*.

$$U = U(Y, X, Q)$$

Individuals are asked to pay a dollar amount, *W*, for access to a specific public land, *Q*. The utility after paying this amount would be

$$U_1 = U_1(Y - W, X, Q) + \epsilon_1$$

while the utility if the bid is rejected would be

$$U_0 = U_0(Y, X) + \epsilon_0$$

Individuals will pay *W*, if $U_1 \geq U_0$. That is,

$$\Pr[\text{Yes}] = \Pr[U_1(Y - W, X, Q) + \epsilon_1 \geq U_0(Y, X) + \epsilon_0$$

which can be rewritten as

$$\Pr[\epsilon_0 - \epsilon_1] \leq U_1(Y - W, X, Q) - U_0(Y, X).$$

Let $\Delta U = U_1(Y - W, X, Q) - U_0(Y, X)$, $\eta = \Delta_0 - \Delta_1$, and $F\eta$ be the cumulative distribution function of the error. Then the expression above can be rewritten as

$$\Pr[(\eta) \leq \Delta U] = F\eta(\Delta U),$$

which, if $F\eta(\Delta U)$ is assumed to have a logistic cumulative density function, is equal to $(1 + e^{-\Delta U})^{-1}$.

Using the approach described by Hanley et al. (1997) in order to proceed, we need to adopt a specific functional form for $u(.)$. Assume, for example, the following simple form:

$$u = u(\alpha + \beta_1 Y + \beta_2 X + \beta_3 Q)$$

Then the change in utility would be

$$\Delta U = U_1(Y - W, X, Q) - U_0(Y, X) = [\alpha_1 + \beta_1^*(Y - W) + \beta_2 X + \beta_3 Q] - [\alpha_2 + \beta_1 Y + \beta_2 X]$$
$$= (\alpha_1 - \alpha_2) - \beta_1 W + \beta_3 Q.$$

Then, the probability of a Yes response is:

$$\Pr[\text{Yes}] = F\eta[(\alpha_1 - \alpha_2) - \beta_1 W + \beta_3 Q].$$

The median WTP is calculated by

$$\Pr[U_1(Y - W, X, Q) > U_0(Y, X)] = 0.5.$$

We will use the approximation of compensating surplus, using the formula derived by Hanemann (1984)

$$\Pr[\text{Yes}] = \left(1 + \mathrm{e}^{-\alpha-\beta W}\right)^{-1}$$

then the median WTP$=-\alpha/\beta$

In a binary regression α is the sum of the coefficients of the explanatory variables, multiplied by the mean value of each variable, and β is the coefficient for the variable representing the bid amount.

The mean WTP is calculated by

$$\text{mean WTP} = \int_0^T \left[1 - G_{\text{wtp}}\right] \mathrm{d}W$$

where G_{wtp} is the distribution function of the true willingness to pay. T is infinite for the true willingness to pay and is truncated at some value for the purpose of estimation.

References

Alberini, A., Boyle, K., Welsh, M., 2003. Analysis of contingent valuation data with multiple bids and response options allowing respondents to express uncertainty. Journal of Environmental Economics and Management 45, 40–62.

Blamey, R., Bennett, J., Morrison, M., 1999. Yea-saying in contingent valuation surveys. Land Economics 75 (1), 126–141.

Brown, T., Champ, P., Bishop, R., McCollum, D., 1996. Which response format reveals the truth about donations to a public good? Land Economics 72 (2), 152–166.

Carson, R.T., Hanemann, W.M., Kopp, R.J., Krosnick, J.A., Mitchell, R.C., Ruud, P.A., Smith, V.K., 1994. "Prospective Interim Lost Use Due to DDT and PCB Contamination in the Southern California Bight." A Report of Natural Resource Damage Assessment, Inc. and Industrial Economics, Inc. to US. NOAA.

Champ, P., Bishop, R., Brown, T., McCollum, D., 1997. Using donation mechanisms to value nonuse benefits from public goods. Journal of Environmental Economics and Management 33 (2), 151–163.

Champ, P., Alberini, A., Correas, I., 2003. "Using Contingent Valuation to Value a Noxious Weeds Control Program: The Effects of Including an "Unsure" Response Category." Working paper.

Cooper, J.C., 1994. A comparison of approaches to calculating confidence intervals for benefit measures from dichotomous choice contingent valuation surveys. Land Economics 70 (1), 111–122 ((February)).

Dillman, D.A., 2001. Mail and Internet Surveys: The Tailored Design Method. John Wiley and Sons, Inc., New York.

Duffeld, J.W., Patterson, D.A., 1991. Inference and optimal design for a welfare measure in dichotomous choice contingent valuation. Land Economics 67 (2), 225–239.

Ekstrand, E.R., Loomis, J., 1997. Estimating Willingness to Pay for Protecting Critical Habitat for Threatened and Endangered Fish With Respondent Uncertainty. Unpublished manuscript.

Hanemann, M., 1984. Welfare evaluations in contingent valuation experiments with discrete responses. American Journal of Agricultural Economics 66 (3), 332–341.

Hanley, N., Shogren, J.F., White, B., 1997. Environmental Economics in Theory and Practice. Oxford University Press, Oxford.

Harrison, G.W., Rustrom, E.E., in press. Experimental evidence on the existence of hypothetical bias in value elicitation methods. In: Plott, C., Smith, V.L. (Eds.), Handbook of Results in Experimental Economics. New York: Elsevier Science.
Holmes, T., Kramer, R., 1995. An independent sample test of yea-saying and starting point bias in dichotomous-choice contingent valuation. Journal of Environmental Economics and Management 29, 121–132.
Kanninen, B., 1995. Bias in discrete response contingent valuation. Journal of Environmental Economics and Management 28, 114–125.
Krinsky, I., Robb, A.L., 1986. On approximating the statistical properties of elasticities. Review of Economics and Statistics 68, 715–719 (November).
Li, C., Mattsson, 1995. Discrete choice under preference uncertainty: an improved structural model for contingent valuation. Journal of Environmental Economics and Management 28, 256–269.
List, J.A., Graig, A., 2001. What experimental protocol influence disparities between actual and hypothetical stated values? Environmental and Resource Economics 20, 241–254.
National Oceanographic and Atmospheric Administration (NOAA), 1994. Proposed rules for valuing environmental damages. Federal Register 59, 1062–1091 (January).
Park, T., Loomis, J.B., Creel, M., 1991. Confidence intervals for evaluating benefits estimates from dichotomous choice contingent valuation studies. Land Economics 67 (1), 64–73 (February).
Polasky, S., Gainutdinova,O., Kerkvliet, J., 1996. Comparing CV responses with voting behavior: open space survey and referendum in Corvallis, Oregon. In Herriges, J. (Compiler). W-133 Benefits and Costs Transfer in Natural Resources Planning: 9th Interim Report. Department of Economics, Iowa State University, Ames, IA.
Ready, R., Navrud, S., 1999. “How Do Respondents With Uncertain Willingness to Pay Answer Contingent Valuation Questions?” Manuscript, Agricultural University of Norway.
Schwarz, N., Sudman, S., 1996. Answering Questions. Methodology for Determining Cognitive and Communicative Processes in Survey Research. Jossey-Bass Publishers, San Francisco.
Wang, H., 1997. Treatment of don’t know responses in contingent valuation surveys: a random valuation model. Journal of Environmental Economics and Management 32 (2), 219–232.
Welsh, M., Bishop, R., 1993. “Multiple Bounded Discrete Choice Models.” In W-133 Benefits and Costs Transfers in Natural Resources Planning. 6th Interim Report. Department of Agricultural Economics, University of Georgia.
Welsh, M.P., Poe, G.L., 1998. Elicitation effects in contingent valuation: comparisons to a multiple bounded discrete choice approach. Journal of Environmental Economics and Management 36, 170–185.