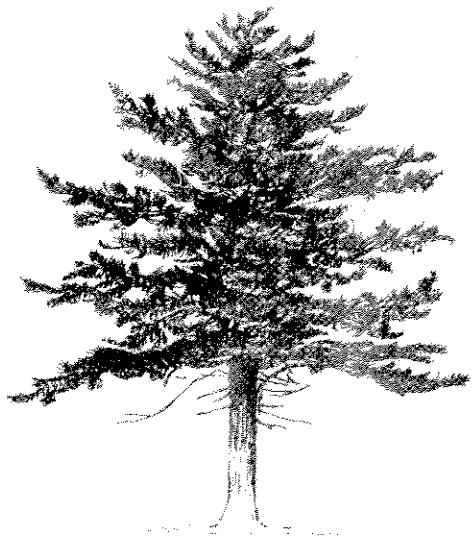
Working Paper Series in

ENVIRONMENTAL & RESOURCE ECONOMICS



A Comparison of Hypothetical Phone and Mail Contingent Valuation Responses for Green Pricing Electricity Programs

Robert G. Ethier, Gregory L. Poe and William Schulze

December 1997

CORNELL UNIVERSITY

ERE 97-06
WP 97-25

A Comparison of Hypothetical Phone and Mail Contingent Valuation Responses for Green Pricing Electricity Programs[1]

Robert Ethier, Gregory Poe, and William Schulze
Cornell University

Jeremy Clark
University of British Columbia

Abstract: To date, much of the policy and research debate on contingent valuation mode effects has relied on experiences drawn from other research disciplines. This study provides the first contingent valuation phone-mail comparison that meets current standards for response rates, draws from a general population, is relevant to the valuation of general environmental goods, and allows comparisons with actual sign-ups. Consistent with previous research in other disciplines, social desirability bias is found in responses to subjective questions -- thus leading to more environmentally favorable responses on the phone. However, this effect does not carry over to hypothetical participation decisions. Hypothetical bias is found in both modes. Yet, application of calibration methods using debriefing questions provided nearly identical values across modes. As such, neither mode appears to dominate from the perspective of providing more valid estimates of actual participation decisions. The selection of survey mode must be based on other criteria.

---

[1] The authors are indebted to Daniel Rondeau, Steven Rose, and Eleanor Smith for their contributions to various components of this research. We wish to thank the National Science Foundation, the U. S. Environmental Protection Agency, and Niagara Mohawk Power Corporation as sponsors of this research. Specifically, we wish to acknowledge Theresa Flaim, Janet Dougherty, Mike Kelleher, Pam Ingersoll, and Maria Ucchino at Niagara Mohawk Power Corporation. Additionally, we thank Pam Rathbun and colleagues at Hagler Bailly, Inc. Madison who also greatly contributed to this study. Funding for this research was provided by NSF/EPA Grant R 824688 under the 1995 Partnership for Environmental Research: Decision Making and Valuation for Environmental Policy Program. Working Paper Series in Environmental and Resource Economics WP 97-06, Department of Agricultural, Resource, and Managerial Economics, Cornell University.

A Comparison of Hypothetical Phone and Mail Contingent Valuation Responses for Green Pricing Electricity Programs

1. Introduction

An unresolved issue in contingent valuation (CV) is whether personal and telephone interviews provide more valid or "better" responses than mail survey techniques. Mail surveys presently dominate CV research because they are easier to implement and cost substantially less than phone interviews, thus enabling a greater amount of survey research to be conducted. However, in an influential review, the National Oceanographic and Atmospheric Administration (NOAA) Panel on contingent valuation maintained that it is "unlikely that reliable estimates of values could be elicited with mail surveys," and recommended in-person or phone interviews [Arrow *et al.* 1993, p. 4611]. In response to the Panel's report, Dillman (1993) argued that problems with personal interviews are generally understated, while the shortcomings of mail surveys are overstated. Others like Bishop *et al.* (1988: p. 336) found that "the data gathered through self-administered surveys may, other things being equal, be better than that obtained by telephone surveys."

Evidence on elicitation "mode" effects is drawn primarily from work by survey researchers in other fields. Unfortunately, little direct empirical evidence is available from CV studies to shed light on this debate. Two exceptions are found in recent CV phone/mail comparisons performed by Whittaker *et al.* (1997) and by Loomis and King (1994). Loomis and King asked for willingness to pay (WTP) for improved wildlife habitat in California. However, they attained relatively low response rates for both the mail and phone portions of their survey and drew samples from slightly different populations, which reduces the comparability of CV responses by increasing the likelihood of differing sample biases. Whittaker *et al.* elicited WTP for increased fees at state parks, and achieved high response rates with a survey from a self-selected sample of Colorado State Park users. Yet, in their own assessment of their work they note that fee increases involve potential strategic issues and that this sample was not drawn from the general population. A further limitation of these two studies, from the perspective of CV validity, is that neither had actual behavior to serve as a

reference for evaluating mode effects. Thus, the critical question of whether phone or mail responses provide better estimates of actual behavior remains unanswered by existing CV research.

This paper offers a unique contribution to the CV survey literature by providing a direct telephone/mail comparison with response rates in the 70% range (as suggested by the NOAA Panel), utilizing a sample drawn from a general population (a sampling strategy corresponding to most CV research), and eliciting WTP for a good with high non-use value (a common CV application). In addition, we are able to compare hypothetical phone and mail results with actual participation in a Niagara Mohawk Power Corporation green electricity pricing program. Thus we can not only test for mode effects, but also for hypothetical bias in CV responses. Calibration of responses across modes is also explored to investigate whether different modes necessitate different levels of correction.

The remainder of the paper is organized as follows. Potential mode related survey errors and biases are briefly reviewed in the next section. Section 3 reviews the existing CV research on mode effects for the purpose of formulating a set of testable research hypotheses. The fourth section summarizes the survey methodology. Results and discussion are provided in Sections 5 and 6.

2. Survey Errors and Possible Mode Effects

The question of the relative validity of mail versus phone surveys is important because of the high cost of phone surveys. Another concern about contingent valuation remains the potential for *hypothetical bias*. Past research has demonstrated that "hypothetical questions, particularly about donations to generally desirable environmental goods, seem to engender overestimates of actual WTP" [Brown *et al.* 1996, p. 164]. A separate body of research suggests that there are mode effects on survey responses, including stated willingness to pay. Yet, because these two concepts have not been linked in a single study, it remains to be determined whether phone or mail elicitation acts to minimize hypothetical bias in valuation, and whether different types of calibration rules are needed across modes.

In addition to this broad concern about hypothetical bias, the survey nature of CV introduces other possible errors in drawing inferences about social values for public goods. *Sampling error* is the first likely source of error. This error may be induced by differences in how the sample is drawn, resulting in a divergence between the survey and the target populations if elements of the target population are systematically excluded from the sample frame. Such divergence has been termed sample frame bias, coverage bias, or non-coverage error [Edwards and Anderson, 1987; Loomis and King; Groves, 1987; Dillman, 1991]. A standard example of this type of bias is that telephone directory listings often used in survey research exclude a substantial portion of the population (perhaps up to 35 to 40 percent or higher) because of unlisted numbers, recent moves, and non-phone households [Schuman, 1996; Traugott, Groves and Lepkowski, 1987]. Thus if alternative modes rely on different sample frames and the sample composition affects estimates of sample values, then mode effects may be observed. Note that this problem results from faulty sampling design, not in how the surveys are implemented.

Another form of sampling error that may occur is *non-response bias*. If the decision to complete and return the survey is systematically related to individual attributes, the resulting sample may not accurately reflect the population being sampled. Inferences about population values may be biased because a non-representative sample would result. If alternative survey modes have different patterns of non-response, then different modes would be expected to exhibit different degrees of this type of bias. Quantification and subsequent weighting of responses to account for sampling effects have been suggested as a technique to correct for non-response [Loomis, 1987; Dalecki *et al.*, 1993]. However, the extent to which respondents answers would differ from those of non-respondents is difficult to quantify and is seldom known. One solution is to keep the proportion of non-respondents as low as possible [Schuman], with the NOAA directives identifying 70% as an appropriate response rate for reducing the likelihood of non-response bias to acceptable levels. At this or similar response rates, "relatively few (mode) differences have been reported for personal attributes", suggesting that reported demographics can be treated as an accurate representation of the sample population [Dillman, 1991, pp. 241-42].

Suppose that the sampling method and high response rates to mail and phone surveys do generate comparable samples as defined by observable "objective" demographic characteristics. It is still possible that different modes will not produce the same answers to non-demographic questions. This could occur because of *sample selection bias*, where respondents having the same observable demographic characteristics, but different unobservable characteristics, respond to a survey with a different likelihood. Such sample selection bias might be attributed to an unobservable *avidity*, whereby those most interested in an issue are most likely to respond to a survey about that issue in spite of the fact that their observable demographic characteristics do not differ form the rest of the population [Loomis and King; Mitchell and Carson, 1989]. Arguments made by Arrow *et al.* and others suggest that an *avidity effect* is more likely to occur in mail surveys because of the ease of non-response (by the non-avid) relative to telephone surveys. As with non-response bias, high response rates are offered as protection against sample selection bias [Dillman, 1991]. That is, if response rates are high, then the importance of these errors should be (relatively) small.

A third factor that might cause a deviation between phone and mail responses is *measurement error*. Differential measurement errors due to mode effects would result in different answers to the same questions by the same respondent for different survey modes. Schuman and others [Schuman and Presser, 1981; Dillman, 1991] have noted that phone respondents tend to give more extreme answers to survey questions. A prime explanation for this, especially for relatively uncomplicated survey topics, is that some respondents may desire to avoid embarrassment and to project a favorable image to the researchers.[2] Such a *social desirability bias* would be expected to affect, in predictable directions, a wide range of questions of a "subjective" nature which might reflect upon the respondent. For example, Connelly and Brown (1994) show that check-off contributions to wildlife on New York State tax forms are over reported in a mail survey by more than twice actual contributions after adjusting for recall errors. Ketosh and Traugott (1981) find that reported voting

[2] Dillman (1991) discusses the three likely explanations for the presence of mode effects: social desirability bias, context effects, and "pace and control" effects stemming from differences in the answering process. We discuss only social desirability bias in this paper, as the expected effects of context and pace in this case are unclear, and in any case are thought to be less significant.

behavior overstates support for the winning candidate by over 10 percent. The broad survey literature suggests that social desirability bias will be higher for both phone and in-person interviews than for mail surveys [Whittaker *et al.*; Dillman 1978], and that mail surveys produce fewer such response effects [Dillman, 1991] . As an example, Dillman and Tarnai (1991) find that the frequency of socially desirable answers given in phone surveys is significantly higher than those in mail surveys for questions about driving while intoxicated. For CV, evidence of social desirability bias effects on WTP responses has been conjectured for phone and in-person surveys [Arrow *et al.*]. However, this hypothesis has only been supported indirectly by evidence that WTP is affected by different interviewers in in-person interviews [Boyle and Bishop, 1988; Mannesto and Loomis, 1991].

3. Previous Phone/Mail CV Comparisons and Testable Propositions

By necessity, the discussion in the previous section drew primarily from survey researchers in other disciplines. To our knowledge, only two CV studies have explicitly compared phone and mail implementations of the same survey. Loomis and King conducted a survey of WTP for the improvement of wildlife habitat and fisheries resources in California. They used random digit dialing for the phone survey, but supplemented phone listings with auto registrations for the mail survey. Thus their sampling frames differed, introducing the possibility of different sample populations. Using a Chi-square test they found statistically significant differences between respondents for four demographic characteristics: gender, education level, age and income. This result is not particularly surprising in light of their low response rates (35% mail and 55% phone), suggesting a combination of sample non-response and sample-selection bias. In spite of these differences in underlying demographics, Loomis and King did not find that broad attitudes about wildlife differed between survey modes. Nevertheless, estimated WTP distributions from a series of dichotomous choice questions were found to be significantly different between mail and phone surveys for three of five programs tested. Mail surveys produced higher WTP estimates for four of the five hypothetical programs. In one case, the estimated mail WTP value was more than triple the estimated phone value. Loomis and King attributed higher mail WTP values to "sample selection effects in mail

surveys" (p.318), which is consistent with the sample non-response bias and avidity effects[3] discussed above. Thus, the low response rate for their mail survey is the likely source of bias.

The study by Whittaker *et al.* used a split sample design with identical phone and mail instruments to ask Colorado state park users dichotomous choice questions about their WTP increased user fees at Colorado state parks. Note that these fees would be paid by all park users, not just respondents. Drawing from a homogeneous sample across modes, the survey produced high response rates (78% mail, 79% phone), thus minimizing the possibility of sampling error. Such high response rates might be anticipated because their survey population consisted of state park users who previously completed a one page survey and who volunteered their addresses and phone numbers. Statistical tests were conducted to evaluate differences between modes for four variables (number of years visiting Colorado state parks, number of parks visited per year, expenditures on recreation equipment, and household income). Only income was significantly different across modes. Since a limited number of fee values were used, a WTP distribution was not estimated. Instead, using a Chi-square test, Whittaker *et al.* found significant differences in WTP for four of nine WTP point estimates, and nearly significant differences for three of the remaining five. In this study, telephone respondents were always more willing than mail respondents to say yes to a fee increase, in one case by a 67% to 43% margin. These findings are opposite those of Loomis and King. Whittaker *et al.* attributed this reversal in part to social desirability bias created by the presence of an interviewer, which is likely to be pronounced because of the sample used and the commodity selected. That is, the conditions were almost ideally designed to induce social desirability effects: respondents were obviously interested in the commodity; and because they had already been contacted in the park, were more likely to perceive that interviewers would be favorable to resource conservation. Whittaker *et al.* also suggest that there may have been strategic bias, in that respondents may seek

[3] An alternative position, suggested by Michael Welsh of Hagler Bailly, Inc., is that social desirability bias might act to lower hypothetical WTP and reduce hypothetical bias. In the presence of a researcher, respondents might take greater efforts to answer truthfully about how much they actually would be willing to pay if they perceive that such consideration and accuracy would please the researcher. In other words, they don't want to be in a position where they might be perceived to be lying to the researcher.

to discourage the park agency from raising fees. They argued that strategic bias would comparatively lower mail results because respondents would have more time to develop strategic thought.

In all, the research literature on mail and phone surveys in other disciplines, as well as the specific CV work by Loomis and King and Whittaker *et al.* suggest that there may be important, systematic differences between phone and mail survey techniques. However, the likelihood of sampling errors in the Loomis and King study, and the possible self-selection effects in the Whittaker *et al.* study, potentially limit the relevance of these studies to broader CV research. Using samples with identical coverage, drawn from general populations, our research further explores some of these issues by examining five propositions about mode effects and hypothetical bias.

Proposition 1: *Phone and mail surveys do not differ in sample response bias as evidenced in observable respondent demographics.* Loomis and King found significantly different demographic values across modes for four variables, while Whittaker *et al.* found differences only in income. Using sample demographics and Chi-square tests of differences in distributions, we test for sample non-response bias across modes. If sample demographics are not determined to be significantly different between phone and mail surveys, we conclude that non-response bias does not differ across modes. Such a result would allow us to attribute any observed mode differences in WTP to alternate explanations (i.e. avidity, social desirability, and/or hypothetical bias).

Proposition 2: *Phone and mail surveys do not yield different responses to non-WTP questions.* As discussed in the previous section, avidity effects are more likely to occur in mail surveys because of the ease of non-response, while social desirability bias is thought to be more prevalent in a phone survey where the presence of an interviewer may influence respondent answers. While we are unable to separate avidity and social desirability effects in this research, we can compare responses to non-WTP questions to gauge the relative effects of such biases. If the green pricing program is thought to be a socially responsible or approved program, avidity and social desirability bias would work in the same direction but on different survey modes. Chi-square tests of difference in distributions are used to determine if response patterns to four questions likely to be inflated by either bias are

significantly different between phone and mail surveys. If the patterns do differ, a larger mean response for the mail survey would support the conclusion that avidity effects outweigh any social desirability effect, while a larger phone survey mean value would suggest that social desirability bias dominates.

Proposition 3: *Phone and mail surveys do not lead to different hypothetical participation decisions.* Whittaker *et al.* concluded that the use of phone surveys results in higher willingness to say 'yes' to a hypothetical fee increase for all park users, which might be caused by social desirability bias not present in a mail survey. In contrast, Loomis and King found that mail surveys exhibited higher WTP than phone elicitations. Given these conflicting results, the null hypothesis is that phone and mail results are the same. A Chi-square test of proportions is used to test equality between unconditional sign-up rates.

Proposition 4: *Phone and mail surveys do not differ in the direction and degree of hypothetical bias.* Even if mode effects do occur, it is not apparent which method is preferred. Alternative methods might systematically overstate, understate, or bracket "true" WTP. In our comparison, however, we make use of actual participation rates obtained with a provision point elicitation mechanism. This mechanism has been shown to approximate demand revelation in the laboratory (see Poe *et al.* 1997 and Rose *et al.* 1996, 1997), and thus provides an appropriate reference for comparing the direction and the degree of hypothetical bias associated with alternative elicitation techniques. Specifically, we use tests of proportions to compare hypothetical and actual "yes" responses to dichotomous choice questions of willingness to participate in a green electricity pricing program. If actual sign-ups differ from hypothetical sign-ups, we conclude that there is hypothetical bias. Combining these results with the results from Proposition 3 allows us to draw conclusions about hypothetical bias effects across modes.

Finally, if hypothetical bias is found to exist with both phone and mail surveys, we consider the following proposition regarding calibration.

Proposition 5: *Phone and mail participation estimates demonstrating hypothetical bias may be corrected with identical calibration.* If hypothetical bias is found to exist, the more interesting and useful question is whether similar calibration methods provide similar participation estimates. Champ *et al.* (1995, 1997), have suggested a calibration question to be placed immediately after the dichotomous choice WTP question. In this approach, participants who respond "yes" to the hypothetical $6/month question are asked how certain (on a scale from 1 to 10, where 1 is very uncertain and 10 is very certain), they are that they would sign up for the program if it were indeed offered. Champ *et al.* found that including only those respondents with a certainty level of ten provided results comparable to actual participation levels. They used a voluntary contributions mechanism for collecting actual contributions, however, which has been demonstrated to under-reveal demand. When a demand revealing provision point mechanism was used instead to collect actual contributions, Poe *et al.* found that a certainty level of seven or higher was the appropriate calibration factor. Here, comparability of calibrated values is tested using difference of proportions tests.

4. Survey Methodology

In 1995-1996 the Niagara Mohawk Power Corporation (NMPC), a public utility in New York State, launched Green Choice™, the largest program in the country for the green pricing of electricity [Holt and Associates, 1997]. NMPC's 1.4 million households were offered the opportunity to fund a green electricity program that would invest in renewable energy projects (e.g. landfill gas reclamation, wind power) as substitutes for traditional energy sources, and a tree planting program. In recent years such green pricing programs have generated substantial interest as utilities come under increasing pressure to provide alternative sources of electricity for customers who prefer environmentally friendly energy sources. Although windpower stations in Traverse City, Michigan and Fort Collins, Colorado have been funded through voluntary surcharges on electricity bills, comparisons of market research and actual sign-up rates for a range of green pricing programs generally suggest that the high levels of interest predicted from marketing research have been seriously biased. While marketing research suggests that 'majorities' of customers will sign-up for

such programs, sign-up rates have generally been below 2 percent [Brynes *et al.*, 1995; Holt and Associates].

Building on the mechanism design recommended by Schulze (1994), NMPC's GreenChoice™ provision mechanism incorporated three key features: a provision point, a money back guarantee, and extended benefits. This provision point mechanism works as follows. If, for example, 12,000 customers joined the GreenChoice™ program in the first year at a fixed $6/month (collected as a surcharge on their electricity bill), $864,000 would be collected and the provision point would be met. NMPC would then plant 50,000 trees and fund a landfill gas project which could replace fossil fuel generated electricity for 1,200 homes. However, if after one year Green Choice™ participation was insufficient to reach the provision point, NMPC would cancel the program and refund all the money that was collected. If money were collected in excess of the provision point, additional moneys would be put toward increasing the scope of the program (extended benefits). The characteristics of the program itself were based on prior market research for NMPC [Wood *et al.*, 1994]. The demand revelation characteristics of the program's funding mechanism are further discussed in an experimental context in Rose *et al.* (1997).

Provision point, money back guarantee mechanisms have been shown to improve participation levels in public programs [Bagnoli and McKee 1991, Rondeau *et al.* 1996, Schulze 1994], and to approximate aggregate demand in 'large group' laboratory experiments [Rondeau *et al.* 1997]. Having a demand revealing mechanism is an important component of validity testing. As argued in Poe *et al.*, Navrud and Veisten (1997), and Foster *et al.* (1997), previous public goods CV validity tests that compared hypothetical and "simulated" markets (e.g. Duffield and Patterson, 1992; Seip and Strand, 1992; Brown *et al.*; and Champ *et al.* 1997) have used voluntary contribution mechanisms which have been demonstrated to under reveal 'true' demand, thus providing biased estimates of hypothetical bias. Because the provision point mechanism has been demonstrated to more closely approximate demand by reducing free-riding, it should better reflect actual preferences. In turn, this should allow for more accurate determination of hypothetical bias and provide an appropriate reference for calibration.

The survey instruments used in this study followed the Dillman Total Design Method [Dillman, 1978). The survey was pretested by administering successive draft versions by phone until respondents clearly understood the instrument. A split-sample was used to compare three designs: actual sign-ups obtained through phone solicitation, and phone and mail responses to hypothetical versions of the actual solicitation question. Initial sample sizes were 250 households for actual sign-ups, and 500 households each for hypothetical phone and mail surveys. Established multiple contact survey techniques, including a two dollar incentive, were used in all versions with Cornell University as the primary correspondent. All three survey versions were administered by Hagler-Bailly Inc. The subject population was NMPC customers in the greater Buffalo, New York metropolitan area. Sample frames were obtained from Genesys, Inc., a private marketing firm, which used identical methods to draw the samples.

In each survey version, respondents were first screened to establish that they were NMPC customers and to determine their previous knowledge of the GreenChoice™ program. A description of the GreenChoice™ program followed, with questions to aid the respondents' understanding. The program description followed the NMPC Green Choice™ brochure as closely as possible and emphasized various components of the good (trees and renewable energy) and the provision point mechanism. The description was followed by a dichotomous choice question asking the respondents if they would sign up for the program. The survey concluded with demographic questions.

Following the dichotomous choice calibration method suggested in Champ *et al.*, participants in the hypothetical surveys were first asked if they would be willing to participate in the program if it was offered at $6/month, and, if they answered "yes", were next asked how certain they were of actually participating in the program on a scale from '1' ("Very Uncertain") to '10' ("Very Certain"). Participants in the actual phone survey were offered the opportunity to sign up for the program at $6/month, with the charge to appear on their monthly bill. This sign-up now/pay later approach follows standard green pricing methods [Holt and Associates]. Furthermore, the phone solicitation

approach corresponded with the "keep it simple" approach adopted by NMPC, which allowed such phone sign-ups for the actual program.[4]

One of the motivations for using mail surveys is their ease of implementation and their relatively low cost per respondent compared to a phone survey.[5] In this study the final cost per respondent varied by nearly a factor of 2 between the phone and mail treatments: the final cost per telephone respondent was estimated to be $45.95, while the final cost per mail respondent was estimated to be $24.12. Thus the price difference between phone and mail surveys is indeed large. Thus it is likely that ruling out the use of mail surveys could curtail both the use of CV generally by increasing implementation costs, and also the quality of CV surveys as sample sizes may be truncated to fit fixed budgets. In contrast, if mail and phone surveys are found to produce comparable results, cheaper mail survey techniques could be used.

5. Survey Results and Discussion

After subtracting disconnected phone numbers, households that had moved, non-NMPC customers, etc., the final response rates were 70.4% for the actual phone survey, 71.2% for the hypothetical phone survey, and 68.3% for the hypothetical mail survey. None of these rates are significantly different at the 10 percent level in pairwise comparisons. The final number of contacts for each were 145, 275, and 299, respectively. The proportion of respondents to the actual survey who chose to join the Green Choice™ program at $6/month was 20.4%. The comparable proportions for the hypothetical surveys were 30.6% for the phone sample and 35.7% for the mail survey. These results are summarized in Table 1.

---

[4] From the perspective of CV validity, testing the lack of price points other than $6/month is a limitation. A standard feature of dichotomous choice surveys is to use different price levels for different subsamples, allowing a WTP function or distribution to be estimated. Because the actual Green Choice program was only offered at $6/month, our comparisons are limited to this point estimate of WTP. As discussed in Balistreri *et al.* (1996), a single point estimate does not allow a separation of spread and location effects on mean WTP. Nevertheless, such comparisons do provide essential insights into the validity of CV [Cummings *et al.*, 1995].

[5] In other circumstances, a mail survey might permit the use of visual aids or allow the researcher to convey more complex information than is practical over the phone. Since the same surveys were conducted over the phone and through the mail, with minor wording changes to make them amenable to each mode, specialized visual aids were not utilized.

Table 1. Survey Results

	Actual Phone Sign-Ups	Hypothetical Phone Sign-Ups	Hypothetical Mail Sign-Ups
Response Rate	71.4%	71.2%	68.3%
Number Answering Sign-Up Question	142	258	286
Participation Rate	20.4%	30.6%	35.7%

As an aside, the number of usable surveys, i.e. returned surveys minus surveys with item non-response, was similar for the hypothetical phone and mail surveys: 85.5% of phone surveys and 83.8% of mail surveys received had non-missing responses to all 10 demographic and other variables used in subsequent analyses. These proportions are not significantly different, suggesting that item non-response is similar between mail and phone surveys.

We now turn to statistical evaluation of the propositions.

Proposition 1: *Phone and mail surveys do not differ in sample response bias as evidenced in observable respondent demographics.* Although overall response rates were not significantly different across modes, an interesting divergence in response patterns occurs in the results obtained from the first screening question. It asks whether survey respondents are NMPC customers. The proportion of respondents saying 'no' to this question varied greatly between hypothetical phone and mail surveys.[6] The proportion of phone respondents contacted who said that they were not NMPC

[6] Note: for simplicity, except for WTP questions, we will restrict our comparisons to hypothetical phone and hypothetical mail treatments. Actual phone response patterns did not differ from hypothetical phone responses.

customers or had moved (a very small percentage) was 11.4%, which is significantly different (χ^2=20.56) from the 3.8% reported by mail customers. We hypothesize that two effects might be occurring: 1) NMPC customers are using the screening question as an excuse to avoid the phone survey, and/or 2) non-NMPC customers fail to return the mail survey and thus are counted as non-respondents. In either case, there is some evidence of mode effects on response. If indeed the case was that phone customers overstated their non-NMPC percentage, the response rate would be 67.9% for the phone survey. If instead the mail non-NMPC percentage was understated because of non-response, the mail survey response rate would be 72.4%. Under either assumption, the response rates would still be comparable across modes.

Consistent with the statistically similar response rates, evidence of response effects is not found in the demographic variables. The distributions of five demographic variables were compared across hypothetical phone and mail surveys using Chi-square tests: age, gender, income, occupation, and level of schooling. None was significantly different at the 10% level of certainty[7] (see Table 2). Thus, there is no observable difference in sample demographics, suggesting similar samples. This result contrasts with both the Loomis and King and Whittaker *et al.* studies, which found differences in their respondent demographics. Because there is no evidence of differences in sample demographics across modes in our survey, any differences in survey responses across modes found while examining the following propositions will be attributed to other mode effects and not sample non-response bias.

Proposition 2: *Phone and mail surveys do not yield different responses to non-WTP questions.* Four questions which might be expected to capture both an avidity effect or social desirability bias were asked: "In the last two years, have you contributed to any environmental groups...?", "Please rate the overall service you receive from NMPC?", "How interested are you in the goal of planting trees

[7] Note that two sets of categories (out of 11) were combined for occupation, one set because of small expected values (<5) and another (sales and services) because their substitutability had seemingly led to coding inconsistencies. The sales/service ratios were 6/19 for phone and 14/2 for mail, which contrasted markedly with the remaining eight categories. Since mail was self-coded and phone was coded using the first selected answer, inconsistencies were likely in similar categories like sales and service.

Table 2. Demographic Variables: Differences Between Phone and Mail Surveys

Variable	Chi-square (n)	df	Phone Mean	Mail Mean
Age	5.18 (553)	6	52.32	53.55
Gender	.863 (564)	1	47.8% male	51.7% male
Income	4.25 (512)	5	$35,747	$39,096
Occupation	11.58 (561)	8	n.a.	n.a.
School	9.65 (560)	7	13.7 years	14.0 years

*,**,*** Significant at .10, .05, and .01, respectively.

on public lands in upstate New York?", and "How interested are you in renewable energy?" Chi-square test results for the four questions are shown in Table 3. We found differences in three of the four "subjective" questions asked in each survey. Differences in 'Rate service' and 'Plant trees' are significant at the 1% level, while 'Give to environment' is significant at the 10% level. For all four questions evaluated, the phone survey produced higher mean responses than the mail survey. This suggests that social desirability bias present in the phone survey outweighs any avidity effects which might have been captured in the mail survey. Such a finding is consistent with Dillman (1991), as well as with the conclusions of Whittaker *et al.*, but contrasts with Loomis and King, who may have had an artificially high avidity effect associated with sample non-response bias. Thus, in this aspect our results are consistent with previous survey research in other disciplines.

Table 3. Subjective Variables: Differences Between Phone and Mail Surveys

Variable	Chi-square (n)	df	Phone Mean	Mail Mean
			-- Percent --	
Give to Environmental Causes	3.09* (561)	1	22.68	16.78
Service (1-10)	20.55*** (561)	6	8.45	7.96
Renewables (1-10)	11.09 (533)	9	6.31	6.16
Trees (1-10)	17.86*** (539)	6	8.42	7.75

*,**,*** Significant at .10, .05, and .01, respectively.

Proposition 3: *Phone and mail surveys do not lead to different hypothetical participation decisions.* The results of this hypothesis are subject to interpretation. On one hand the hypothetical mail (35.7%) sign-up rate was about 17 percent higher than the hypothetical phone (30.6%) sign-up rate, a difference that is arguably large. Yet, unlike previous studies, this difference is not significantly different at the 10% level ($\chi^2 = 1.589 < \chi^2_{.10} = 2.71$), even though the sample size used in these comparisons exceeds the single bid value samples typically used in dichotomous choice CV studies. Note also that the direction of the observed difference between the hypothetical results is the opposite of what one would expect if the phone answers were influenced by social desirability bias, as was found for Proposition 2. Given this divergence, we hypothesize that the nature of the WTP question, with the inclusion of a dollar amount, works to reduce the effect of social desirability bias as found in Proposition 2 by 'grounding' the respondent in a way which does not occur in an 'opinion' question. Thus, social desirability bias might be present in non-WTP type questions in a manner consistent with previous research in other disciplines, but without carryover to valuation questions.

These results are the opposite of Whittaker *et al.*, but consistent with Loomis and King. One means of reconciling our results with those of Whittaker *et al.* is to emphasize their secondary explanation for low mail WTP levels: strategic respondents trying to avoid the possibility of an actual fee increase. Since our hypothetical WTP question involves only the voluntary, fixed price payment by the respondent, strategic incentives to influence actual policies by under-revealing WTP are limited [Hoehn and Randall, 1987]. Indeed, the reverse strategy might be likely: respondents might seek to convince NMPC that green power is highly valued and thus ought to be provided. In addition, the Whittaker *et al.* sample population may have been optimally designed to induce social desirability bias in the phone interview. Regardless of cause, the WTP results are not found to be significantly different. We thus conclude that hypothetical phone and mail WTP results are comparable.

Proposition 4: *Phone and mail surveys differ in the direction and degree of hypothetical bias.* A Chi-square test on the proportions of respondents signing up for the GreenChoice™ program showed that the actual sign-up rate of 20.4% was indeed significantly different from either of the hypothetical sign-ups. Both the actual-hypothetical phone difference ($\chi^2 = 4.837$) and the actual-hypothetical mail difference ($\chi^2 = 10.455$) are significant at the 10% level ($\chi^2_{0.10} = 2.71$). Both hypothetical questions overestimate actual contributions. Therefore, if hypothetical dichotomous choice responses are to be used to predict actual enrollment, some sort of calibration will be necessary.

Proposition 5: *Phone and mail participation estimates demonstrating hypothetical bias may be corrected with identical calibration*. While differences in hypothetical phone and mail willingness to participate responses may be subject to interpretation, calibration of these hypothetical responses produced remarkably similar phone and mail WTP patterns (see Table 4). None of the differences across calibrated values is statistically significant. This result is important. It suggests that while mode effects might affect WTP responses, calibration of these responses demonstrate similar valuation patterns. Since the objective is to predict "true" WTP, it appears that the alternative modes do not provide different estimates of WTP once they are calibrated.

Table 4. Calibrated Hypothetical Sign-Ups and Actual Sign-Ups

	Actual Phone Sign-Ups	Hypothetical Phone Sign-Ups	Hypothetical Mail Sign-Ups
		-- Percent --	
Participation Rate	20.4	30.6	35.7
Certainty ≥ 5*	--	29.1	32.2
Certainty ≥ 6	--	24.8	24.1
Certainty ≥ 7	--	20.9	21.7
Certainty ≥ 8	--	14.0	15.6
Certainty ≥ 9	--	8.5	8.8
Certainty ≥ 10	--	6.6	6.1

*Sign-ups adjusted for certainty are counted as 'yes' if the respondent answered 'yes' to the dichotomous choice question at $6 per month and said that she would actually sign-up with a certainty level greater than or equal to X.

In this analysis, calibration levels of seven or greater for each mode were nearly identical to the actual sign-up rate of 20.4%: hypothetical phone sign-ups at a certainty level of seven or greater were 20.9%, while mail sign-ups at a certainty level greater than or equal to seven were 21.7%. This convergence suggests that a certainty level of seven or greater is an effective way of approximating actual WTP from a hypothetical survey. It is important to note that these results differ from those of Champ *et al.*, who found that a certainty level of '10' corresponded to actual sign-ups. It is possible that calibration levels may vary with commodity. Nonetheless, we argue that a likely explanation for the difference between calibration levels is our use of a demand revealing mechanism. Recall that the voluntary contribution mechanism used to collect actual payments in the Champ *et al.* studies under-reveals true willingness to pay, and thus may lead to over-calibration associated with a higher level of certainty.

6. Discussion

With the increased application of CV there has been substantial debate over whether CV surveys should be conducted by mail, phone or in person interviews. Unfortunately most of the arguments for and against different models largely rely on experiences from survey research conducted in other disciplines. To date, little empirical CV research on the relationship between mode and WTP has been conducted. And, the few studies that have attempted to address this issue provide results that may have limited relevance to broad CV research because of their specific design features.

In contrast to previous CV efforts to examine mode effects on survey response, this survey achieved high response rates for an environmental good with a high nonuse component, and used identical sampling techniques to draw samples from a general population. This is also the first study of mode effects to compare different hypothetical methods to an actual willingness to participate.

Using a green electricity pricing program as a case study, a comparison of hypothetical mail and phone sign-up rates found no difference in demographic variables when survey response rates approximated the NOAA threshold of 70%. Thus, measurable non-response bias does not appear to be a problem when response rates approach acceptable levels. In spite of the similar demographics, mode effects *are* observed in responses to subjective, non-WTP 'opinion' questions. Notably, social desirability bias seems to affect phone answers, resulting in higher responses to what might be regarded as socially desirable 'opinion' questions. This result is consistent with survey research conclusions drawn from other disciplines. However, such social desirability effects attributed to phone elicitation do not appear to carry over to the WTP question. Indeed, the proportion of hypothetical sign-ups was found to be lower by phone than by mail, but not significantly so.

Consistent with previous CV validity research, hypothetical bias was found for dichotomous choice questions, with both hypothetical mail and phone survey responses significantly

overstating actual participation levels obtained through a phone sign-up program. An important finding of this research is that calibration questions based on certainty of "yes" responses provided statistically identical patterns across modes. This finding suggests that if calibration techniques are to be used to adjust hypothetical response to more closely reflect actual willingness to pay, then the calibrated responses are independent of survey mode.

In all, some evidence of social desirability and mode effects does appear across modes, but these effects do not carry over to phone and mail estimates of willingness to participate, whether calibrated or uncalibrated. As such, neither mode appears to dominate from the perspective of providing more valid estimates of actual participation decisions. We therefore argue that the selection of survey mode must be based on other criteria.

References

Arrow, K., R. Solow, E. Leamer, P. Portney, R. Randner, and H. Schuman, 1993. "Report of the NOAA Panel on Contingent Valuation." *Federal Register* 58 (10): 4601-4614.

Bagnoli, M. and M. McKee, 1991. "Voluntary Contributions Games: Efficient Private Provision of Public Goods." *Economic Inquiry* 29: 351-366.

Balistreri, E., G. McClelland, G. Poe, and W. Schulze, 1996. "Can Hypothetical Questions Reveal True Values? A Laboratory Comparison of Dichotomous Choice and Open Ended Contingent Valuation with Auction Values." *Working Paper Series in Environmental and Resource Economics 96-01*, Cornell University, Department of Agricultural, Resource, and Managerial Economics.

Bishop, G., H. Hippler, N. Schwarz, and F. Strack, 1988. "A Comparison of Response Effects in Self-Administered and Telephone Surveys." In *Telephone Survey Methodology*, R. Groves, P. Biemer, L. Lyberg, J. Massey, W. Nicholls, and J. Waksberg, eds. John Wiley & Sons, New York.

Boyle, K. J. and R. C. Bishop, 1988. "Welfare Measurement Comparisons." *American Journal of Agricultural Economics* 70 (1). February 1988: 20-28.

Brown, T. C., P. A. Champ, R. C. Bishop, and D. W. McCollum, 1996. "Which Response Format Reveals the Truth About Donations to a Public Good?", *Land Economics*, 72(2: May): 152-166.

Brynes, B., M. Rahimzadeh, K. Baugh and C. Jones, 1995. "Caution Renewable Energy Fog Ahead! Shedding Light on the Marketability of Renewables." Paper presented at the NARUC-DOE Conference on Renewable and Sustainable Energy Strategies in a Competitive Market, Madison, WI. May 1995.

Champ, P.A., R.C. Bishop, T.C. Brown, and D.W. McCollum, 1995. "A Comparison of Contingent Values and Actual Willingness to Pay Using a Donation Provision Mechanism with Possible Implications for Calibration." In *Benefits and Costs Transfer in Natural Resource Planning,* Western Regional Research Publication W-133, 8: 387-410.

—, 1997. "Using Donation Mechanisms to Value Nonuse Benefits from Public Goods." *Journal of Environmental Economics and Management,* 33(June): 151-162.

Connelly, N. A. and T. L. Brown, 1994. "Effect of Social Desirability Bias and Memory Recall on Reported Contributions to a Wildlife Income Tax Checkoff Program." *Leisure Sciences,* 16: 81-91.

Cummings, R. G., G. W. Harrison, and Rutström, 1995. "Homegrown Values and Hypothetical Surveys: Is the Dichotomous Choice Approach Incentive Compatible?", *American Economic Review,* 85(1): 260-266.

Dalecki, M. GT., J. C. Whitehead, and G. C. Blomquist, 1993. "Sample Non-Response Bias and Aggregate Benefits in Contingent Valuation: An Examination of Early, Late, and Non-Respondents", *Journal of Environmental Management*, 38: 133-143.

Dillman, D.A., 1978. *Mail and Telephone Surveys: The Total Design Method.* John Wiley and Sons, New York.

—, 1991. "The Design and Administration of Mail Surveys." In *Annual Review of Sociology.* Eds. W.R. Scott and J. Blake, eds. Palo Alto, CA.

—, 1993. Letter to the Office of Policy, Planning and Evaluation, U.S. EPA, reprinted in *The Contingent Valuation of Environmental Resources*, D. Bjornstad and J.R. Kahn, eds. Edward Elgar, Brookfield, VT, 1996.

Dillman, D.A. and J. Tarnai, 1991. "Mode Effects of Cognitively Designed Recall Questions: A Comparison of Answers to Telephone and Mail Surveys." In *Measurement Errors in Surveys*. P.P. Biemer, R.M. Groves, L.E. Lyberg, N.A. Mathiowetz, and S. Sudman, eds. John Wiley, New York, 1991.

Duffield, J.W. and D.A. Patterson, 1992. "Field Testing of Existence Values: An Instream Flow Trust Fund for Montana Rivers." Paper presented at the AERE/ASSA meetings, New Orleans, January 1992.

Edwards, S., and G. Anderson, 1987. "Overlooked Biases in Contingent Valuation Surveys." *Land Economics* 63 (2): 168-178.

Foster, V., I.J. Bateman, and D. Harley, 1997. "Real and Hypothetical Willingness to Pay for Environmental Preservation: A Non-Experimental Comparison." *Journal of Agricultural Economics* 48(2): 123-138.

Groves, R. M., 1987. "Research on Survey Data Quality", *Public Opinion Quarterly*, 51: s156-s172.

Hoehn, J. P. and A. Randall, 1987. "A Satisfactory Benefit-Cost Estimator from Contingent Valuation", *Journal of Environmental Economics and Management*, 14: 226-247.

—, 1997. Green Pricing Newsletter. The Regulatory Assistance Project, Gardiner, ME. Vol. 5, May.

Katosh, J. P., and M. W. Traugoff, 1981. "The Consequences of Validated and Self-Reported Voting Measures", *Public Opinion Quarterly*, 45: 519-535.

Loomis, J., 1987. "Expanding Contingent Value Sample Estimates to Aggregate Benefit Estimates: Current Practices and Proposed Solutions." *Land Economics* 63: 396-402.

——— and M. King, 1994. "Comparison of Mail and Telephone-Mail Contingent Valuation Surveys." *Journal of Environmental Management* 41: 309-324.

Mannesto, G. and J. B. Loomis, 1991. "Evaluation of Mail and In-Person Contingent Valuation Surveys: Results from a Study of Recreational Boaters", *Journal of Environmental Management*, 32: 177-190.

Mitchell, R. and R. Carson, 1989. *Using Surveys to Value Public Goods: The Contingent Valuation Method*. Resources for the Future, Washington, D.C.

Navrud, S. and K. Veisten, 1997. "Using Contingent Valuation and Actual Donations to Bound the True Willingness-to-Pay, Unpublished Manuscript, Department of Economics and Social Sciences and Department of Forestry, Agricultural University of Norway.

Poe, G.L., J. Clark, and W. D. Schulze, 1997. "Can Hypothetical Questions Predict Actual Participation in Public Programs: A Field Validity Test Using a Provision Point Mechanism." Cornell University *Working Paper Series in Environmental and Resource Economics 97-05*, Department of Agricultural, Resource and Managerial Economics. Paper presented at the European Association of Environmental and Resource Economists, June 1997, Tillburg.

Rondeau, D., G.L. Poe, and W.D. Schulze, 1996. "Developing a Demand Revealing Market Criterion for Contingent Valuation Validity Tests." *Working Paper 96-16*, Department of Agricultural, Resource, and Managerial Economics, Cornell University. Paper presented at the annual meetings of the American Agricultural Economics Association, Aug. 1997, Toronto.

Rondeau, D., W. D. Schulze, and G. L. Poe, 1997. "Voluntary Revelation of the Demand for Public Goods Using a Provision Point Mechanism", Cornell University *Working Paper Series in Environmental and Resource Economics 97-04* , Department of Agricultural, Resource, and Managerial Economics.

Rose, S., J. Clark, G. Poe, D. Rondeau, and W. Schulze, 1996. "Field and Laboratory Tests of a Provision Point Mechanism." Paper presented at the Economic Science Association Meeting, October 1996, Tucson, AZ.

Rose, S., J. Clark, G. Poe, D. Rondeau, and W. Schulze, 1997. "The Private Provision of Public Goods: Tests of a Provision Point Mechanism for Funding Green Power Programs." *Working Paper Series in Environmental and Resource Economics 97-02*, Department of Agricultural, Resource, and Managerial Economics, Cornell University. Paper presented at the annual meetings of the Northeastern Agricultural and Resource Economics Association, June 1997, Sturbridge, MA.

Schulze, W.D., 1994. "Green Pricing: Solutions for the Potential Free Rider Problem." Paper prepared for Niagara Mohawk Power Corporation.

Schuman, H., "The Sensitivity of CV Outcomes to CV Survey Methods", *The Contingent Valuation of Environmental Resources*, D. Bjornstad and J.R. Kahn, eds. Edward Elgar, Brookfield, VT, 1996.

Schuman, H. and S. Presser, 1981. *Questions and Answers in Attitude Survey: Experiments on Question Form, Wording, and Context.* Academic Press, San Diego.

Seip, K. and J. Strand, 1992. "Willingness to Pay for Environmental Goods in Norway: A Contingent Valuation Study with Real Payment." *Environmental and Resource Economics* 2: 91-106.

Traugott, M., R. Groves, and J. Lepkowski, 1987. "Using Dual Frame Designs to Reduce Nonresponse in Telephone Surveys", *Public Opinion Quarterly*, 51: 522-439.

Whittaker, D., J. Vaske, M. Donnelly, and D. DeRuiter, 1997. "Mail Versus Telephone Surveys: Potential Biases in Expenditure and Willingness-To-Pay Data." Under review at the *Journal of Leisure Research.*

Wood, L.L., W.H. Desvousges, A.E. Kenyon, M.V. Bala, F.R. Johnson, R. Iachan, and E.E. Fries, 1994. "Evaluating the Market for 'Green' Products: WTP Results and Market Penetration Forecasts." *Working Paper #4,* Center for Economics Research, Research Triangle Institute.

OTHER A.R.M.E. WORKING PAPERS

WP No	Title	Author(s)
97-24	An Exploration of the Differences in National Growth: Some Implications for Future Environmental Policy	Khanna, N., T.D. Mount and D. Chapman
97-23	A Critical Overview of the Economic Structure of Integrated Assessment Models of Climate Change	Khanna, N. and D. Chapman
97-22	Valuation of Groundwater Quality: Contingent Values, Public Policy Needs, and Damage Functions	Poe, G.L.
97-21	Can Hypothetical Questions Predict Actual Participation in Public Programs? A Field Validity Test Using a Provision Point Mechanism	Poe, G., J. Clark and W. Schulze
97-20	Voluntary Revelation of the Demand for Public Goods Using a Provision Point Mechanism	Rondeau, D., W.D. Schulze and G.L. Poe
97-19	The Bioeconomics of Marine Sanctuaries	Conrad, J.M.
97-18	Introducing Recursive Partitioning to Agricultural Credit Scoring	Novak, M. and E.L. LaDue
97-17	Trust in Japanese Interfirm Relations: Institutional Sanctions Matter	Hagen, J.M. and S. Choe
97-16	Effective Incentives and Chickpea Competitiveness in India	Rao, K. and S. Kyle
97-15	Can Hypothetical Questions Reveal True Values? A Laboratory Comparison of Dichotomous Choice and Open-Ended Contingent Values with Auction Values	Balistreri, E., G. McClelland, G. Poe and W. Schulze
97-14	Global Hunger: The Methodologies Underlying the Official Statistics	Poleman, T.T.
97-13	Agriculture in the Republic of Karakalpakstan and Khorezm Oblast of Uzbekistan	Kyle, S. and P. Chabot
97-12	Crop Budgets for the Western Region of Uzbekistan	Chabot, P. and S. Kyle
97-11	Farmer Participation in Reforestation Incentive Programs in Costa Rica	Thacher, T., D.R. Lee and J.W. Schelhas
97-10	Ecotourism Demand and Differential Pricing of National Park Entrance Fees in Costa Rica	Chase, L.C., D.R. Lee, W.D. Schulze and D.J. Anderson

To order single copies of ARME publications, write to: Publications, Department of Agricultural, Resource, and Managerial Economics, Warren Hall, Cornell University, Ithaca, NY 14853-7801.