Was the NOAA Panel Correct about Contingent Valuation?

Richard T. Carson, W. Michael Hanemann, Raymond J. Kopp, Jon A. Krosnick, Robert C. Mitchell, Stanley Presser, Paul A. Ruud, and V. Kerry Smith

with Michael Conaway and Kerry Martin

Discussion Paper 96-20

May 1996

Resources for the Future
1616 P Street, NW
Washington, DC 20036
Telephone 202-328-5000
Fax 202-939-3460

Was the NOAA Panel Correct about Contingent Valuation

Richard T. Carson, W. Michael Hanemann, Raymond J. Kopp,
Jon A. Krosnick, Robert C. Mitchell, Stanley Presser,
Paul A. Ruud, and V. Kerry Smith

with

Michael Conaway and Kerry Martin

Abstract

The past few years have seen a highly charged debate about whether contingent valuation (CV) surveys can provide valid economic measures of people's values for environmental resources. In an effort to appraise the validity of CV measures of economic value, a distinguished panel of social scientists, chaired by two Nobel laureates, was established by NOAA, to critically evaluate the validity of CV measures of nonuse value.

The Panel provided an extensive set of guidelines for CV survey construction, administration, and analysis, and distinguished a subset of items from their guidelines for special emphasis and described them as *burden of proof* requirements. Of particular interest was the Panel's requirement that CV surveys demonstrate "responsiveness to the scope of the environmental insult." That demonstration has come to be called a *scope* test. The paper reports the findings from the first CV study that adheres to the NOAA Panel's guidelines and includes a formal scope test.

Key Words: contingent valuation, scope test, NOAA Panel

JEL Classification No(s).: D6, H4

Table of Contents

List of Tables

Was the NOAA Panel Correct about Contingent Valuation?

Richard T. Carson, W. Michael Hanemann, Raymond J. Kopp, Jon A. Krosnick,
Robert C. Mitchell, Stanley Presser, Paul A. Ruud, and V. Kerry Smith

with

Michael Conaway and Kerry Martin*

I. INTRODUCTION

Seven years ago, in March 1989, the Exxon Valdez ran into the submerged rocks of Bligh Reef and released 11 million gallons of crude oil. The intervening years have seen a highly charged debate about whether contingent valuation (CV) surveys can provide valid economic measures of people's values for environmental resources that they may never actively use.[1] In an effort to appraise the validity of CV measures of economic value, the Bush

---

* The authors are, respectively: Associate Professor of Economics, University of California (San Diego); Professor of Agricultural and Natural Resource Economics, University of California (Berkeley); Senior Fellow, Resources for the Future; Associate Professor of Psychology and Political Science, Ohio State University; Professor of Geography, Clark University; Professor of Sociology, University of Maryland (College Park); Professor of Economics, University of California (Berkeley); and Arts and Sciences Professor, Duke University, and University Fellow, Resources for the Future. Conaway is currently and Martin was associated with Natural Resource Damage Assessment, Inc. at the time this research was undertaken. Both made extensive contributions throughout the effort. The work described in this paper was funded by the Damage Assessment Center of the National Oceanic and Atmospheric Administration as part of a natural resource damage assessment under contract number 50-DGNC-1-00007. Additional support to aid in the preparation of this paper was provided to Smith by the UNC Sea Grant Program under Grant No. R/MRD-25. All opinions expressed in this paper are those of the authors and should not be attributed to the National Oceanic and Atmospheric Administration, the UNC Sea Grant Program, or any persons or organizations acknowledged above.

[1] The debate initiated with a conference in Washington, DC in April 1992 sponsored by the Exxon Corporation. The results of some of the research sponsored by Exxon are reported in Hausman [1992]. Other CV related work supported by Exxon has recently been published in economics and law journals, see McFadden [1994] and Boyle et al. [1994] as examples. Results of the Federally supported research associated with the Exxon case were never released. The research supported by Alaska has circulated in an unpublished report [Carson et al., 1992] and in an unpublished discussion paper [see Carson et al., 1995].

The research reported here is the first attempt (to our knowledge) to draw together public research conducted as part of natural resource damage assessments to address specific issues that bear on the reliability of CV. Diamond and Hausman [1994], for example, note in evaluating contingent valuation research that:

> We have argued that internal consistency tests [scope and adding up] (particularly adding up tests) are required to assess the reliability and validity of such surveys. When these tests have been done,

Administration's General Counsel for the National Oceanic and Atmospheric Administration (NOAA), Thomas Campbell, appointed in 1992 a distinguished panel of social scientists, chaired by two Nobel laureates, Kenneth Arrow and Robert Solow, to critically evaluate the validity of CV measures of nonuse value.[2,3] The evaluation was conducted within the specific context of the assessment of natural resource damages due to releases of hazardous substances or oil into the environment, but has general applicability to the use of CV.[4]

The Panel provided an extensive set of guidelines for CV survey construction, administration, and analysis. In the Panel's view, " . . . the more closely the guidelines are followed, the more reliable the result will be" [Arrow et al., 1993, p. 4609]. In addition, the Panel distinguished a subset of items from their guidelines for special emphasis and described them as *burden of proof* requirements. In describing the elements with this special focus, the Panel stated:

> ". . . if a CV survey suffered from any of the following maladies, we would judge its findings 'unreliable':
>
> - a high nonresponse rate to the entire survey or to the valuation question

---

contingent valuation comes up short. Contingent valuation proponents typically claim that the surveys used for these tests were not done well enough. Yet they have not subjected their own surveys to such tests (p.62, bracketed terms and emphasis added).

[2] The Panel included, in addition to Arrow and Solow, Edward Leamer, Paul Portney, Ray Radner and Howard Schuman. Their report, Arrow et al. [1993], was published in the *Federal Register* Jan. 15, 1993, 4601-4614.

[3] To better understand the context for the NOAA Panel, it is useful to recognize the long history of contingent valuation. The method was first proposed in 1947 and its first reported application was by Davis [1963] in his Harvard Ph.D. dissertation on the economic value of recreation in the Maine woods. Numerous applications of the method to various public goods and studies of its methodological properties were conducted in the 1970's and 1980's both in the United States and, increasingly, in other countries. A review of the theoretical and empirical basis of contingent valuation at the end of this period is presented in Mitchell and Carson [1989]. A recent contingent valuation bibliography [Carson et al., 1994] contains over 1600 references to books, articles, and reports on the method. Brief histories of CV can be found in Portney [1994] and Hanemann [1994].

[4] For discussion of the background for the NOAA panel see Portney [1994] and Campbell [1993].

- inadequate responsiveness to the scope of the environmental insult
- lack of understanding of the task by the respondents
- lack of belief in the full restoration scenario
- 'yes' or 'no' votes on the hypothetical referendums that are not followed up or explained by making reference to the cost and/or the value of the program"

[Arrow et al., 1993, p. 4609]

The second item in this list, "inadequate responsiveness to the scope of the environmental insult," has attracted the most attention and is regarded by many as an acid test for CV studies. The demonstration that CV results are responsive to different levels of the environmental insult has come to be called a *scope* test.

Given these guidelines and burden of proof requirements, the Arrow-Solow Panel concluded its report noting that:

> . . under those conditions (and others specified above), CV studies convey useful information. We think it is fair to describe such information as reliable by the standards that seem to be implicit in similar contexts, like market analysis for new and innovative products and the assessment of other damages normally allowed in court proceedings.
>
> ...CV [contingent valuation] produces estimates reliable enough to be the starting point of a judicial process of damage assessment, including passive - use values [i.e., nonuse values]. [5]

While there has been extensive debate over the Panel's conclusion and its guidelines, none of the CV studies considered by the Panel had conducted a survey following the Panel's recommended guidelines or performed a test of scope.[6]

---

[5] Arrow et al. [1993] p. 4610, bracketed phrase inserted.

[6] For a critical review of the NOAA panel report see Cummings and Harrison [1995].

The purpose of this paper is to summarize the findings from several recent contingent valuation studies conducted by the authors and to report the findings from the first study that adhered to the NOAA Panel's guidelines. Overall, our results indicate that stated choices (i.e., choices in response to CV questions) do reveal smaller willingness to pay (WTP) for smaller amounts of an environmental commodity provided by a specifically described program or plan and, therefore, meet the "scope test" of the burden of proof requirement. Of course, this does not prove that CV results are, in general, valid and reliable. Rather, it offers the first evidence that the NOAA Panel guidelines do inform the evaluation process and, more importantly, it demonstrates that measures of stated choices, assembled from information conforming to the Panel's guidelines, do display the type of responsiveness to economic, demographic, and attitudinal factors observed with revealed preference measures of people's economic values.

In addition, we describe the findings from a coordinated program of study that was structured to address CV design questions posed by the Panel as requiring further research. Since these unanswered questions bear on CV design, they had to be addressed prior to the design and execution of the full study described below. The design questions concern the effects of: the timing of the survey in relationship to the events giving rise to natural resource injuries (e.g., oil spills); the role of a "would-not-vote" option in a referendum style CV question; and the impact of interviewer effects and what the panel described as "social desirability bias" for in-person, referendum style CV surveys.

II. BACKGROUND, NOAA PANEL GUIDELINES, AND SURVEY DESIGN

In what follows, we summarize selected results from four CV surveys, all based on in-person interviews with a combined total of over 5,000 observations. Table 1 describes the object of choice,[7] sample frame, focus group, pre-test, and pilot activities as well as sample characteristics for each survey. Two surveys, the original survey undertaken as part of the State of Alaska's natural resource damage assessment for the Exxon Valdez oil spill and our replication of that survey in 1993 (labeled NORC, for the National Opinion Research Center), consider the same object of choice -- a plan to prevent future oil spills. The other two surveys comprise the base and scope instruments used for the assessment of natural resource injuries due to PCB and DDT contamination in the waters off Southern California.[8] Because our objective is to evaluate the Arrow-Solow Panel's proposed methodology, we will not describe all the details associated with the specific elements of the natural resource injuries presented in each of the three survey instruments used in our research program.

---

[7] Economic values are constructed for *objects of choice* under *particular circumstances of choice*. From the analyst's perspective, the object of choice is the thing for which an economic value is desired. Objects of choice may be quite general and extend far beyond our normal conception of private goods sold in markets. Objects of choice can be public goods like local police protection, ambient air and water quality, or species and habitat protection. But the list does not stop here. Objects of choice can be any tangible or intangible object, process or activity that can be described in a way that allows a choice to be fashioned.

[8] A more complete discussion of the design of these two surveys is available in Carson et al. [1994]. *Base* and *scope* are terms used to identify the two survey instruments, where the magnitude of the injuries described to respondents was smaller in the scope instrument than the base instrument. Copies of the base and scope questionnaires are available on request.

The survey questionnaire was designed to estimate prospective interim lost use value (where use value was defined to conform to the definition offered by the Court of Appeals to include passive use values) for losses due to the injuries to natural resources caused by DDT and PCB's released into the South Coast, an area near Los Angeles, lying within and along the northern part of the Southern California Bight. Four species in this area were described as impacted by the DDT and PCB's in the sediment in the area -- two of birds -- Peregrine Falcons and Bald Eagles and two of fish -- White Croaker and Kelp Bass.

Table 1: Features of the Surveys

Attribute	Alaska	NORC	Southern California[a] Base	Southern California[a] Scope
Date of Survey	1991	1993	1994	1994
Population	U.S.	12 PSU's[b]	California	California
Object of Choice	Plan to provide two Coast Guard ships to escort oil tankers in Prince William Sound to prevent future accidents and avoid future injuries due to oil spills	Same as Alaska	Plan to accelerate recovery of reproduction problems of four species by 45 years[c]	Plan to accelerate recovery of reproduction of two species by 10 years[d]
Sample Size	1,043	1,182	1,857	953
Nature of Payment	One-time addition to Federal Income Taxes	One-time addition to Federal Income Taxes	One-time addition to California State Income taxes	One-time addition to California State Income taxes
Tax Amounts	$10, $30 $60, $120	$10, $30 $60, $120	$10, $25, $80 $140, $215	$10, $25, $80 $140, $215
Focus Groups	7	-	5	9
Pre-Tests	2 (12, 18)	2[e] (64, 26)	4 (44, 57, 49, 116)	4 (44, 54, 40, 44)
Pilots	4 (105, 195, 244, 176)	-	4 (332, 460, 324, 473)	-
Response Rate	75.2%	73.0%	72.1%	73.8%

a Sample was intended to represent the population of English speaking Californians, age 18 or older, living in private residences they own or rent (or whose rent or mortgage they contribute to). Thirteen primary sampling units were selected with probabilities proportional to their 1990 Census population counts, including: Del Norte and Humboldt; El Dorado, Placer, Sacramento, and Yolo; Alameda, San Mateo, San Francisco, Marin, and Contra Costa; San Joaquin; Santa Clara; Fresno; Santa Barbara; Ventura; Los Angeles County; Los Angeles City; Orange; Riverside and San Bernardino; and, San Diego. Within the selected PSUs, 652 segments (city blocks, groups of blocks, or Census equivalents in rural areas) were selected with probabilities proportional to their 1990 Census counts of housing units.

b The 12 PSUs selected from NORC's master area probability sample were: Baltimore, MD; Birmingham, AL; Boston, MA; Charleston, SC; Harrisburg, PA; Ft. Wayne, IN; Manchester, NY; Nicholas County, KY; Portland, OR; Richmond, VA; Seattle, WA; and Tampa, FL.

c The four species include two birds: Bald Eagles, Peregrine Falcons; and two fish: White Croaker and Kelp Bass. See Carson et al. [1994] for the description of the injuries.

d For the scope scenario only the two fish species were described as injured.

e These pretests were conducted to evaluate the instructions used with the design variations e.g. ballot box, and would-not-vote and composite versions of the questionnaire, not to evaluate framing.

Monetary measures of economic value are constructed from information about the choices individuals make. When an individual decides to acquire (or support a process leading to) a specific object of choice, we know by his or her agreement to incur a monetary cost that the object obtained is worth at least this cost to this particular individual. The cost provides a lower bound on the individual's willingness-to-pay (WTP). If the object of choice is refused, then the cost isolates an upper bound for WTP. Two aspects of the circumstances surrounding this choice are important: (a) the presence of a recognizable relationship (to the people making the decision) between the object of choice and the monetary cost and (b) the financial consequence of the decision (i.e., incurring the cost). Most discussions of CV focus on the second issue. Their concern is about whether a stated choice with a proposed (but not realized) financial consequence provides the same information as an actual choice. However, the first issue is equally important. Indirect approaches (often referred to as revealed preference approaches) for recovering information about people's values for environmental resources rely on assumed linkages between the choices that can be observed and those resources. A CV survey, by contrast, offers the opportunity to make this link explicit through the plan described for changing the resource.[9]

Each of the four surveys summarized in Table 1 describes a specific plan to undertake actions intended to affect one or more environmental resources. As a result, the object of

---

[9] The intention is to describe plans that closely resemble what could be the actual practice of delivering the intended object of choice. In fact, in the case of the Alaska survey, a plan was adopted that closely resembled what was described in the survey and it had the intended effect of avoiding future spills. A March 20, 1993 *Los Angeles Times* article on oil spills noted that the closest call since the Exxon Valdez was in October 1992 when a tanker, Kenai, had problems with its steering system but avoided pulling up on a rock due to the presence of an escort tug, similar to the one described in the plan proposed in the Alaska survey.

choice is a change in the environmental resource arising through the activities described in a specific program or plan. This design element focuses respondents' attention on a "bundling" of the proposed change in the resource(s) with the plan. It provides a mechanism that connects the object of choice to the payment in a format, judged through qualitative research conducted as part of the CV design, to be plausible and legitimate.[10] It is through this plausibility and legitimacy that CV practitioners seek to insure that each respondent takes seriously the implied financial obligation associated with his or her choice.

In addition to a brief sketch of the objects of choice, Table 1 describes the target population and a few salient features of the research design. The development of each questionnaire involved focus groups, cognitive interviews, and pre-tests in an effort to improve the respondents' understanding of the choices they were asked to make. Pilot surveys were conducted prior to the main survey to evaluate the field performance of the instrument, to evaluate the design selected for the tax amounts (a referenda format was used to elicit values), and to test specific hypotheses bearing on the final instrument design. Four pilots were conducted for the Alaska survey and the base version of the Southern California surveys. For example, one Alaska pilot survey investigated different payment vehicles (i.e., income taxes versus higher oil prices). In the Southern California survey, a split sample pilot investigated the effect of the timing of natural recovery. Another considered the impact on WTP of one less

---

[10] The design stage of the CV surveys developed each choice question as part of a sequence of qualitative and quantitative activities including: multiple focus groups, one-on-one cognitive interviews, pre-tests, and pilot studies. These efforts helped the CV designers understand how the plan should be described, the issues involved in interpreting specific wording as well as the assessment of the field performance of each version of the survey instrument. This extensive development process also helped to assure that respondents understood the object of choice as intended and connected the proposed payment for the plan to that object.

species, when a potentially important bird species, Brown Pelicans, was dropped from the list of injured natural resources. While both the Alaska and the Southern California surveys had four pilots, the development process, objectives of each pilot, and time sequencing of activities were quite different. Quantitative and qualitative aspects of these pilots and the pretests document activities consistent with the Arrow-Solow Panel's recommendations, including: required pre-testing of the questionnaires; judgments associated with adopting a conservative design in the wording; pretesting of photographs and visual aids; and initial checks on understanding and acceptance of information associated with each plan serving as the object of choice.

The Arrow-Solow Panel recommendations for reliable CV surveys can be conveniently divided into three groups: general guidelines, guidelines for value elicitation surveys, and goals for value elicitation surveys. The first column of Table 2 presents these three categories of recommendations, reporting the key phrase the Panel used in describing each suggestion. This listing illustrates that there is some overlap in several of the recommendations. The second and third columns in the table report our evaluation of the degree of adherence to these recommendations for the 1991 Alaska survey (and implicitly the 1993 NORC) and the Southern California surveys (both the base and the scope). Six designations are offered: satisfied (S); not-satisfied (N); judgmental evaluation as satisfied (J); tested with the NORC survey (NORC); irrelevant (I); and unable to be evaluated due to the study design (designated with -).

The remaining sections of this paper will focus more specifically on items Id, IIg, IIh, and IIj of Table 2 as well as the important burden of proof requirements (IIIg). The other items in Table 2 are resolved largely by inspection of documentation underlying the activities

Table 2: NOAA Panel Guidelines[a]

Survey Guidelines			Survey[b]	
			Alaska	Southern California Bight
I.	General			
	a.	Sample Size and Type	S	S
	b.	Minimize Non-Response	S	S
	c.	Personal Interview	S	S
	d.	Pretesting for Interviewer Effects	NORC	NORC
	e.	Reporting	S	S
	f.	Careful Pretesting of CV Questionnaire	S	S
II.	Value Elicitation Surveys			
	a.	Conservative Design	J	J
	b.	Elicitation Format	S	S
	c.	Accurate Description of Program or Policy	J	J
	d.	Pretesting of Photographs	J	S
	e.	Reminder of Undamaged Substitutes Commodities	J	S
	f.	Adequate Time Lapse from the Accident	J	S
	g.	Temporal Averaging	NORC	NORC
	h.	No Answer Option (Would-not-vote)	NORC	NORC
	i.	Yes/No Followups	S	S
	j.	Cross Tabulations	S	S
	k.	Checks on Understanding and Acceptance	S	S
III.	Goals for Value Elicitation Surveys			
	a.	Alternative Expenditure Possibilities	S	S
	b.	Deflection of Transaction Value	-	J
	c.	Steady State or Interim Losses	-	I
	d.	Present Value Calculation of Interim Losses	-	J
	e.	Advance Approval	I	I
	f.	Reliable Reference Surveys	I	I
	g.	Burden of Proof	-	S

[a] The source for the guidelines is the NOAA Panel report, see Arrow et al. [1993] p. 4608-4609.

[b] Six designations are offered -- satisfied (S), not-satisfied (N), judgmental evaluation as satisfied (J), tested with the NORC survey (NORC), irrelevant (I), and unable to be evaluated due to the study design (designated with -).

summarized in Table 1, the questionnaires, and the detailed reports describing the questionnaire design, sampling procedures, verbatim records, and interviewer responses [see Carson et al., 1992; and Carson et al., 1994]. All the surveys involved personal interviews using professional interviewers, conducted by two of the leading survey research groups in the US -- Westat of Rockville, MD for the Alaska and Southern California surveys, and, as noted, the National Opinion Research Center at the University of Chicago for the NORC study.

The NORC study was undertaken to examine the three issues identified earlier as unresolved questions raised in the Arrow-Solow report that would bear on CV survey design. This study consists of four separate surveys based on the earlier Alaska questionnaire, including: (1) a complete replication the original Alaska instrument modified only slightly to reflect the timing of the new interviews in relation to the Exxon Valdez oil spill (termed the *replication* version); (2) a version in which the respondent votes on a paper ballot that is placed in a sealed box and the interviewer does not know the decision (the *ballot box* version); (3) a version where the respondent is told there are three options -- "for," "against," and "would-not-vote" (the *no-vote* version); and (4) a version with the three choice options and the ballot box (the *no-vote/ballot box* version).

Following the Panel's recommendation, a referendum style, stated choice CV question was used in all surveys. All surveys explicitly mentioned the alternative uses of the resources and the proposed cost of the plan prior to eliciting the stated choice for each plan. Each survey asked respondents follow-up questions after their vote was recorded (IIi.). In all surveys, respondents voting for the plan were given opportunities to reconsider their votes and change from a yes to a no, and were also asked if they felt the interview's descriptions and

process "pushed" them to vote in a particular way. In the Southern California survey, specific attention was given to the other areas of Southern California where the four injured species (two fish -- White Croaker and Kelp Bass -- and two birds -- Bald Eagles and Peregrine Falcons) could be found (IIe.).

Table 3: NORC Tests

Issue	Procedure	Tax Amount	χ^2	P-value
Temporal Stability (IIg.)	Compare 1991 Alaska survey with base version of 1993 NORC survey	$10	0.005	0.946(NR)
		$30	0.406	0.524(NR)
		$60	0.036	0.850(NR)
		$120	0.019	0.890(NR)
Would-not-vote Option (IIh.)	Compare base 1993 NORC Survey with version III offering would-not-vote option	$10	0.581	0.446(NR)
		$30	0.662	0.416(NR)
		$60	0.269	0.604(NR)
		$120	0.459	0.498(NR)
Interviewer Effects (Id.)	Compare base 1993 NORC Survey with version II using ballot box	$10	2.092	0.148(NR)
		$30	0.003	0.957(NR)
		$60	0.028	0.868(NR)
		$120	0.036	0.849(NR)
Interaction Effects of Would-not-vote and Interview Effects	Compare base 1993 NORC Survey with Version IV using would-not-vote and ballot box	$10	4.854	0.028(R)
		$30	0.768	0.381(NR)
		$60	0.075	0.784(NR)
		$120	0.117	0.732(NR)

[a] NR - cannot reject null hypothesis; R - Reject null hypothesis at most conventional levels for p-value.

III. THE NORC STUDY

The 1993 replication of the Alaska survey examined the temporal reliability of CV based WTP estimates, the introduction of a "would-not-vote" option within a referendum style value elicitation format, and the potential for a social desirability bias. Because the original Alaska questionnaire and tax amounts were unchanged in the NORC survey, it is possible to

test the temporal stability of economic values with simple contingency tables, avoiding the introduction of other maintained assumptions about the functions governing respondents' answers to stated choice questions. Table 3 summarizes the χ^2 tests conducted at each tax amount. These temporal reliability tests compare the original 1991 Alaska survey with the "replication" version of the NORC survey. Two sets of test results compare the "would-not-vote" and "ballot box" versions to the NORC "replication" version. The fourth test considers a composite case where the respondent was given a referendum question with the "would-not-vote" option that called for a response on a paper ballot placed in a box so the interviewer would not know the choice. This case is also compared to the NORC "replication" instrument. Only one of the 16 tests (i.e., 4 versions and each of 4 tax amounts) provides any indication that changes in the timing of the interview or changes in question format had an effect on stated choices. This arises in the composite version of the questionnaire that combines the "ballot box" with the "would-not-vote" form for a tax amount of $10.[11]

On the question of temporal reliability, the comparison of 1993 responses with the original Alaska survey indicates comparable responses even though the 1991 Alaska and 1993 NORC sample were composed of different households. These stated choices reveal estimates of WTP that are not significantly different between the 1991 and the 1993 responses. The 1991 Turnbull lower bound estimate of expected WTP was $52.80 (with an estimated asymptotic standard error of 2.12) while the 1993 results were $52.81 (4.08).[12] After a simple

---

[11] All of the tests rely on two-by-two contingency tables with "would-not-vote" or "don't know" responses treated as choices against the plan (a conservative coding decision recommended by Schuman [1996]).

[12] The Turnbull lower bound mean was developed using the Turnbull [1976] non-parametric estimator of

adjustment for the effects of inflation (a multiple of 1.061 to convert 1991 to 1993 dollars based on the CPI), there is no significant difference in the WTP estimates based on the two sets of CV choices (i.e., Z = 1.16).[13]

Like the test of temporal reliability, the χ^2 tests of the "would-not-vote" option, using a conservative recoding of "would-not-vote" and "not sure" responses as against the plan, and the results with the secret ballot (designed to reduce the likelihood that someone would report support for the program to please the interviewer), strongly support the position that offering a respondent a "would-not-vote" option has no effect on WTP derived from stated choices in CV surveys of the type characterized by the Alaska and NORC questionnaires. This same conclusion can be drawn in the case of the "ballot box" version designed to test potential for

---

distribution functions based on interval censored data. Assuming discrete choice referendum questions with a single take-it-or-leave-it question design, using tax amounts (t_j), allow respondents to be sorted for these amounts into two groups, defining the distribution function.

F_j = Probability (WTP $\leq t_j$)
$1 - F_j$ = Probability (WTP > t_j)

The log-likelihood function underlying the Turnbull is $l = \sum_{j=1}^{K} [N_j \ln(F_j) + Y_j \ln(1 - F_j)]$

where N_j = number of respondents indicating no to t_j
Y_j = number of respondents indicating yes to t_j
K = number of values for t_j

Turnbull demonstrated how these frequencies can be used to develop an estimate of the distribution function. The lower bound mean (LBM) uses this estimate to compute the mean using the lower bounds of each interval as:

$$\begin{aligned} LBM = \; & 0 \text{ Prob } (0 \leq WTP < t_1) \\ & + t_1 \text{ Prob } (t_1 \leq WTP < t_2) + \ldots \\ & + t_{k-1} \text{ Prob } (t_{k-1} \leq WTP < t_k) \\ & + t_k (1 - F(t_k)) \end{aligned}$$

It was first proposed in Carson et al. [1994]. For an introduction to the method with illustrations see Haab and McConnell [1995].

[13] The test of temporal reliability is discussed more fully in Carson et al. [1995].

social desirability bias with in-person interviews. While the composite version of the questionnaire, combining the ballot box and the would-not-vote options suggests that a difference in voting patterns for the $10 tax amount, this difference is not large enough to lead to a significantly different estimate for the Turnbull lower bound mean WTP from the interaction version. The estimate with the composite version is $50.61 (4.05).

The last component of our test involved the sensitivity of the original specification of the choice equation used in the 1991 Alaska study to describe the influence of economic, demographic, attitude, information, and plan related variables on the stated choices. These models offer a means to evaluate the construct validity of CV responses. As such, it is important to consider whether these issues in question design affected conclusions about construct validity. Overall these test results suggest no significant differences. [14] Based on the analyses described above, the final questionnaire design for the Southern California CV study did not offer a "would-not-vote" option and assumed that there would be no social desirability bias.

---

[14] Multinomial logit models were also estimated to test the implicit restrictions on the effects of economic (e.g. tax amount and income), demographic, and attitudinal variables due to recording "would-not-vote" responses as against the plan. The null hypothesis of equal effects of these factors could not be rejected. (A detailed summary of these results is provided in Carson et al. [1996]). Moreover, multivariate models estimated using respondents' choices in the original sample and the NORC replication treatment did not, for the most part, identify significant differences in the factors influential to stated choices. Several different models, including different estimators (i.e. probit, single-bounded survival, and double-bounded survival), and treatments for the repeated sample were considered. Both multivariate probit and survival model estimates corresponding to the specification of the survival model reported in Carson et al. [1992] with a dummy variable identifying the new survey in 1993 indicated no significant difference between the two samples. A second specification with the dummy for the new sample interacted with all specified determinants also suggested no difference in the effects of most determinants. Only the effect of a variable indicating concerns about coastal oil spills was found to have a significant difference with probit and the simple (single choice) survival model. Thus, these results generally support the conclusion that the choices and the conditioning effects of the principal determinants remained stable between the 1991 and 1993 samples. A detailed summary of these results is provided in Carson et al. [1996].

Table 4: Tests for Scope With Southern California Survey

A. Contingency Tests

Tax Amount		For	Against	χ^2	p-value[a]
$10	Base (%)	209 (55.9)	165 (44.1)	21.50	0.001 (R)
	Scope	72 (35.6)	130 (64.4)		
$25	Base	163 (46.3)	189 (53.7)	23.50	0.001 (R)
	Scope	45 (24.7)	137 (75.3)		
$80	Base	120 (32.9)	245 (67.1)	14.39	0.001 (R)
	Scope	35 (17.9)	161 (82.1)		
$140	Base	102 (26.5)	283 (73.5)	10.00	0.002 (R)
	Scope	29 (14.9)	166 (85.1)		
$215	Base	85 (22.3)	296 (77.7)	10.85	0.001 (R)
	Scope	19 (10.7)	159 (89.3)		

B. Turnbull Lower Bound Mean (LBM)

	LBM	Std. dev.	
Base	$63.24	2.54	
Scope	$34.02	2.82	
Z-test	7.17	(reject null hypothesis of equality of LBM p-value <.001)	
Likelihood Ratio	83.46	(reject null hypothesis of equality of distributions p-value <.001)	

C. Survival Model Test Results

(1) Weibull	
Z-test for Location parameter	(reject null hypothesis p-value <.001)
Likelihood ratio test for Location and scale parameter	(reject null hypothesis p-value <.001)
(2) Log Normal	
Z-test for Location parameter	(reject null hypothesis p-value <.001)
Likelihood ratio test for Location and scale parameters	(reject null hypothesis p-value <.001)

[a] NR - cannot reject null hypothesis; R - Reject null hypothesis at most conventional levels for p-value.

IV. BURDEN OF PROOF EVALUATION FOR THE SOUTHERN CALIFORNIA CV SURVEY

Scope

One of the most important aspect of the Arrow-Solow Panel's *burden of proof* requirements is the proposed test for scope. This test requires a demonstration that stated choices for different amounts of the environmental resource(s) have different implied monetary values. As Table 1 suggests, this test was an integral component of the design of the Southern California study. In the base survey, four animal species -- two birds (Bald Eagles, Peregrine Falcons) as well as two fish (White Croaker and Kelp Bass) -- were described as having reproductive problems in the South Coast area and respondents were told that these species would recover naturally from these problems in 50 years. The survey testing for sensitivity to scope identified only the two species of fish (White Croaker and Kelp Bass) as having reproductive problems and suggested natural recovery would take 15 years. In each case, an accelerated recovery plan was offered that would take 5 years to eliminate the source of these problems (PCB and DDT contamination off the coast). Thus, the object of choice in the base survey was a plan to mitigate the reproductive problems of four species within 5 instead of 50 years and in the scope survey to accomplish this objective for two species, in 5 instead of 15 years. Economic theory would suggest that the monetary measure of the larger object of choice (base survey) should exceed that for the smaller one (scope survey).

Three separate tests of the responsiveness of stated choice estimates of WTP to the scope of the injuries presented in the Southern California survey were undertaken. The least restrictive tests are the simple contingency tests comparing the stated choices in the base and scope samples. Table 4 presents these tests at each of the tax amounts used in the Southern

California survey.[15] There is unambiguous support for a difference in stated choices between the two samples. Moreover, the direction of the difference, with more votes for the plan in the base inquiry description, reveals a greater monetized value for the larger base than the smaller scope object of choice.

This conclusion is confirmed by the results of the two additional tests, also reported in Table 4. The first compares the Turnbull lower bound mean for WTP in both the base and scope samples. Both a simple test for differences in these means and a likelihood ratio test for differences in the non-parametric estimates of the distribution confirm significant differences in the base and scope samples. Simply put, people are willing to pay more for more significant programs. These differences were consistent with our *a priori* expectations.[16]

It is also possible to evaluate whether, following their stated "votes," the respondents in the base and scope samples reported different perceptions of the seriousness of the injuries in each case. This difference is important because at the time of their vote, the respondents in each sample were not aware of the alternative (larger or smaller) injury description. This evaluation of severity was based on an attitude question asked after the CV choice. For the base questionnaire it was:

> All things considered, would you say the *fish and bird* reproduction problems I told you about in the South Coast were not serious at all, not too serious, somewhat serious, very serious, or extremely serious?

---

[15] The contingency table tests were performed on the respondent voting patterns after each respondent who voted for the plan had an opportunity to reconsider (change) his or her vote. These same tests have also been conducted using the respondent voting patterns prior to the reconsideration of their vote. This second test yields the same conclusion as the test reported in Table 5.

[16] We fit a variety of parametric survival functions to the responses. The last component of Table 4 reports the results using the Weibull and Log-Normal models. All the distributions we considered support the overall conclusion of responsiveness in monetary measures of economic value to scope.

For the scope questionnaire this question replaced *fish and bird* with *fish*. A simple two-way chi square test indicated a significant difference in stated seriousness of the problem (i.e. χ^2 = 148.90 with a p value <0.001).[17]

Understanding of Task, Belief in the Plan, and Support for CV Stated Choices as Economic Choices

The remaining components of the Panel's burden of proof requirements include: lack of understanding of the task by the respondents; lack of belief in the full restoration scenario; and 'yes' or 'no' votes on the hypothetical referendums that are not followed up or explained by making reference to the cost and/or the value of the program. Agreement with these requirements is more difficult to assess than scope or response rates because the reference point (what constitutes agreement) is not made clear in the Panel's report. However, there is information in CV surveys that can be used to address the Panel's concerns.

The Panel requires that respondents display an understanding of the task (i.e., the choice to be made) and believe that the restoration plan will lead to full restoration of the injured resources. To collect information responsive to this requirement, open-ended questions were incorporated in the interview. Interviewers were instructed to record respondents' answers as completely as possible on the questionnaire. These responses are referred to as verbatims. During the first part of the interview describing the accelerated recovery plan and injuries, respondents were periodically asked if they wanted material repeated. Those who responded "yes" were asked to describe what they would like repeated. These responses offer information

---

[17] In addition, the scope survey included an additional question at the end of the interview asking respondents if they would consider the reproduction problems more serious if they impacted bald eagles and peregrine falcons in addition to the fish. Seventy-four percent responded yes to this question.

about respondents' reactions and understanding. After hearing descriptions of the bird and fish reproduction problems, respondents were asked if they wanted information repeated. 96 percent answered "no." Of the 62 respondents answering "yes," most wanted information about aspects of the situation that were covered later in the interview.

After the description of the accelerated recovery plan, respondents were asked if they had questions about how it would work, as well as (in a separate question) if they wanted more information about either the accelerated recovery program or natural recovery. About 14 percent of respondents (257) asked about the accelerated recovery plan and most asked about the cost or how the program would work. Approximately the same proportion asked questions about either the plan or the natural recovery option at this prompt (260 respondents). Over 25 percent of those asking at the first prompt (i.e. for the program) did so again. Here, the responses were also consistent with interpreting the material in a meaningful way.

The Panel's burden of proof requires that follow-up questions be asked to assess reasons for a respondent's stated choice. Immediately following the choice question, respondents were asked why they voted for or against the plan.[18] Table 5 summarizes the distribution of verbatim responses for those voting for and against the plan. The majority of those respondents favoring the plan cite reasons related to the accelerated recovery of the injured species. Similarly, those opposing the plan identify other priorities or cost of the program.

---

[18] It is useful to note that this recommendation of the Arrow-Solow Panel reflected the established practice of using such follow-up questions during the survey design and development stage as discussed in Carson et al. [1992].

Table 5: Reasons for Stated Choices:
Southern California Base Case Accelerated Restoration Plan

A. Reasons for the Plan (Question: Can you tell me what covering the contaminated sediments would do that made you willing to pay for it?)

	Category	Percent [n = 907][a]
(a)	Help affected species and/or the area where they live	71.9%
(b)	Hasten recovery process	21.7%
(c)	Respondents personally concerned about environment/ wildlife or perceives household would benefit in some way	16.4%
(d)	Prevent possible physical harm to respondent or others	13.9%
(e)	Feel responsible to help fix this problem	13.6%
(f)	Others such as grandchildren living in area would benefit	12.0%
(g)	Cost affordable/reasonable	9.8%
(h)	Might help other animal/ecosystem	7.5%
(i)	Protect environment	1.9%
(j)	Other	15.6%

B. Reasons to be Against Plan (Question: Did you vote against the program because it wasn't worth that much money to you or because it would be somewhat difficult for your household to pay that much, or because of some other reason?)

	Category	Percent [n = 825]
(a)	Problem not important/other problems more important	51.5%
(b)	Some difficult to pay/cost too high	26.3%
(c)	Concerns about program or payment plan design	21.2%
(d)	Isn't worth that much money	12.5%
(e)	Wants more information	2.1%
(f)	Other	9.3%

[a] Percentages add to more than 100 percent due to multiple responses..

While not part of the Arrow-Solow Panel burden of proof, it seems natural to require that stated choices be responsive to economic, demographic, and attitudinal variables in ways that agree with *a priori* expectations (as one would expect to find with revealed preference responses). Table 6 reports a simple probit model that evaluates whether stated choices (votes from the base survey) are linked to respondents' characteristics and attitudes in predictable ways.

Table 6: Construct Validity Model for Base Injury of Southern California Sample: Probit Estimates

Variable	Coefficient (Z statistic)	p-value	Variable Mean
Intercept	-1.16 (-1.86)	0.063	—
Log of tax amount	-.40 (-12.77)	0.000	4.06
Log of household income if < median California income (35,173); zero otherwise	.17 (2.86)	0.004	5.08
Log of household income if ≥ 35,173 and < 150,000; zero otherwise	.15 (2.70)	0.007	4.85
Log of household income if ≥ 150,000; zero otherwise	.11 (2.17)	0.030	0.45
College associate degree or higher = 1; zero otherwise	-.18 (-2.26)	0.024	0.38
Did not pay California state income taxes = 1; zero otherwise	.49 (3.43)	0.000	0.11
Protecting coastal areas from oil spills extremely important = 1; zero otherwise	.15 (2.00)	0.046	0.37
Protecting coastal areas from oil spills not important = 1; zero otherwise	-.71 (-1.38)	0.169	0.01
Increase spending for endangered wildlife = 1; zero otherwise	.42 (5.30)	0.000	0.48
Decrease spending for endangered wildlife = 1; zero otherwise	-.27 (-2.15)	0.032	0.15
Respondent rates self as not at least somewhat strong environmentalist = 1; zero otherwise	-.24 (-3.09)	0.002	0.39
Respondent concludes natural recovery will take a lot more time = 1; zero otherwise	.52 (3.63)	0.000	0.06
Respondent concludes natural recovery will take a lot less time = 1; zero otherwise	-.29 (-2.69)	0.007	0.15
Respondent expects plan to be completely or mostly effective = 1; zero otherwise	.60 (7.46)	0.000	0.63
Respondent expects plan to be not too effective or not at all effective = 1; zero otherwise	-1.26 (-4.77)	0.000	0.09

Table 6 (continued)

Variable	Coefficient (Z statistic)	p-value	Variable Mean
Respondent asked questions about how plan worked or its cost = 1; zero otherwise	-.30 (-3.58)	0.000	0.24
Respondent does not think will only have to pay special tax for one year = 1; zero otherwise	-.28 (-3.67)	0.000	0.37
Respondent prefers tax payment vehicle over higher prices = 1; zero otherwise	.39 (5.44)	0.000	0.35
Respondent has great deal of confidence in California State Gov't = 1; zero otherwise	.31 (1.72)	0.085	0.04
Respondent has no confidence in California State Gov't = 1; zero otherwise	-.21 (-2.13)	0.033	0.17
Respondent wants increased spending on only one or on no programs = 1; zero otherwise	-.32 (-3.72)	0.000	0.26
Respondent participates in saltwater boating or fishing or often goes to beach = 1; zero otherwise	.22 (2.91)	0.004	0.59
Respondent is birdwatcher = 1; zero otherwise	.18 (2.41)	0.016	0.41
Respondent often watches TV programs about animals and birds = 1; zero otherwise	.19 (2.52)	0.012	0.44
Respondent's household often eats fish = 1; zero otherwise	.18 (2.45)	0.014	0.41
Respondent lives in Los Angeles or Orange County = 1; zero otherwise	.17 (2.17)	0.030	0.33
Respondent lives north of San Francisco Bay Area = 1; zero otherwise	-.25 (-2.10)	0.036	0.11
Sample size	1857		
Pseudo R^2	0.279		
Log - Likelihood	-879.78		

This analysis provides clear support with all economically relevant variables and consistently significant and plausible estimates, for the variables associated with questions involving attitude and respondents' evaluations of the plan.

V. DISCUSSION

Recent criticisms of contingent valuation estimates of monetary values for environmental resources (including nonuse values) have been characterized as suggesting there may be little hope that CV surveys, applied to situations with appreciable nonuse values, could ever pass economic consistency tests.[19] Our findings for a survey that conforms to the Arrow-Solow Panel's guidelines suggest that these earlier conclusions are incorrect. The estimates reported in this paper adhere to all the proposed consistency requirements laid out by the Panel. All estimates, whether based on stated or revealed preference valuation models, are "captives" of the maintained assumptions used to analyze how choices are used to reconstruct the tradeoffs that underlie all monetary measures of people's preferences.

Like all sources of economic data, survey methods offering CV choices provide information along with noise. The issue before researchers seeking to use the information is whether there are protocols that, when replicated, yield stated choice information that is informative about people's preferences. The Arrow-Solow Panel has taken an important step toward developing these guidelines and the results of our study provide evidence supporting many of the Panel's recommendations. However, our supplementary research using the Alaska Exxon

---

[19] Recall our earlier discussion of the Diamond-Hausman conclusions in note 1.

Valdez survey suggests the Panel's concerns about temporal reliability, question format, and social desirability biases appear unwarranted.

Overall, then, based on a series of large scale, in-person surveys, we conclude that there is support for the Arrow-Solow Panel's proposed protocol for CV surveys. While it may be possible to relax these standards, we do not have a basis as yet for determining how adjustments to the recommended practices would influence the economic consistency of the CV estimates of monetary values.

REFERENCES

Arrow, K., R. Solow, P. R. Portney, E. E. Leamer, R. Radner, and H. Schuman. 1993. "Report of the NOAA Panel on Contingent Valuation," *Federal Register*, January 15, vol. 58, no. 10, pp. 4601-4614.

Boyle, Kevin J., William H. Desvousges, F. Reed Johnson, Richard W. Dunford, and Sara P. Hudson. 1994. "An Investigation of Part-Whole Bias in Contingent Valuation Studies," *Journal of Environmental Economics and Management*, vol. 27 (July), pp. 64-83.

Campbell, Thomas A. 1993. "Natural Resource Damage Assessments: A Glance Backward and a Look Forward," *Baylor Law Review*, vol. 45 (Spring), pp. 221-232.

Carson, Richard T., W. Michael Hanemann, Raymond J. Kopp, Jon. A. Krosnick, Robert C. Mitchell, Stanley Presser, Paul A. Ruud, and V. Kerry Smith. 1994. "Prospective Interim Lost Use Value Due to DDT and PCB Contamination in the Southern California," Report to National Oceanic and Atmospheric Administration, Natural Resource Damage Assessment, Inc., La Jolla, Calif., September.

Carson, Richard T., W. Michael Hanemann, Raymond J. Kopp, Jon. A. Krosnick, Robert C. Mitchell, Stanley Presser, Paul A. Ruud, and V. Kerry Smith. 1995a. "Temporal Reliability of Estimates From Contingent Valuation," Discussion Paper 95-37, Resources for the Future, Washington, D.C.

Carson, Richard T., W. Michael Hanemann, Raymond J. Kopp, Jon A. Krosnick, Robert C. Mitchell, Stanley Presser, Paul A. Ruud, and V. Kerry Smith. 1995b. "Referendum Design and Contingent Valuation: The NOAA Panel's No-Vote Recommendation," Discussion Paper 96-05, Resources for the Future, Washington, D.C.

Carson, Richard T., Robert C. Mitchell, W. Michael Hanemann, Raymond J. Kopp, Stanley Presser, and Paul A. Ruud. 1992. "A Contingent Valuation Study of Lost Passive Use Values Resulting From the Exxon Valdez Oil Spill," unpublished report to Attorney General of the State of Alaska (La Jolla, Calif., NRDA, Inc.), November 10.

Carson, Richard T., Robert C. Mitchell, W. Michael Hanemann, Raymond J. Kopp, Stanley Presser, and Paul A. Ruud. 1995. "Contingent Valuation and Lost Passive Use: Damages From The Exxon Valdez," Discussion Paper QE94-18, Resources for the Future, Washington, D.C., revised April 1995.

Carson, Richard T., Nicholas E. Flores, Kerry M. Martin, and Jennifer L. Wright. 1996. "Contingent Valuation and Revealed Preference Methodologies: Comparing the Estimates for Quasi-Public Goods," *Land Economics*, vol. 72 (February), pp. 80-99.

Carson, Richard T., et al. 1994. "A Bibliography of Contingent Valuation Studies and Papers," Natural Resource Damage Assessment, Inc., La Jolla, Calif.

Cummings, Ronald G., and Glenn W. Harrison. 1995. "The Measurement and Decomposition of Non Use Values: A Critical Review," *Environmental and Resource Economics*, vol. 5 (August), pp. 225-247.

Davis, Robert. 1963. "The Value of Outdoor Recreation: An Economic Study of the Maine Woods," doctoral dissertation in economics, Harvard University.

Diamond, Peter. 1996. "Testing the Internal Consistency of Contingent Valuation Surveys," *Journal of Environmental Economics and Management* (forthcoming).

Diamond, Peter A., and Jerry A. Hausman. 1994. "Contingent Valuation: Is Some Number Better Than No Number?", *Journal of Economic Perspectives*, vol. 8 (Fall), pp. 45-64.

Griffin, Charles. C., John Briscoe, Bhanevar Singh, Radhika Ramasubban, and Ramesh Bhatea. 1995. "Contingent Valuation and Actual Behavior: Predicting Connections to New Water Systems in the State of Kerala, India," *World Bank Economic Review*, vol. 9, no. 3, pp. 373-395.

Haab, Timothy C., and Kenneth E. McConnell. 1995. "Referendum Models and Negative Willingness to Pay: Alternative Solutions," unpublished, University of Maryland, April 13.

Hanemann, W. Michael.. 1994. "Valuing The Environment Through Contingent Valuation," *Economic Perspectives*, vol. 8, no. 4, pp. 19-43.

McFadden, Daniel. 1994. "Contingent Valuation and Social Choice," *American Journal of Agricultural Economics*, vol. 76 (November), pp. 689-708.

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

Portney, Paul, R. 1994. "The Contingent Valuation Debate: Why Economists Should Care," *Economic Perspectives*, vol. 8, no. 4, pp. 3-18.

Schuman, Howard. 1996. "The Sensitivity of CV Outcomes to CV Survey Methods," in D.J. Bjornstad and J.R. Kahn, eds., *The Contingent Valuation of Environmental Resources: Methodological Issues and Research Needs* (Aldershot, U.K., Edward Elgar Publishing Co.).

Smith, V. Kerry, and Laura Osborne. 1996. "Do Contingent Valuation Estimates Pass a Scope Test? A Meta Analysis," *Journal of Environmental Economics and Management*, forthcoming.

Turnbull, Bruce W. 1976. "The Empirical Distribution Function with Arbitrarily Grouped, Censored and Truncated Data," *Journal of the Royal Statistical Society*, B38 (No 3), pp. 290-295.