PROSPECTIVE INTERIM LOST USE VALUE DUE TO PCB AND DDT CONTAMINATION IN THE SOUTHERN CALIFORNIA BIGHT

VOLUME 1

SEPTEMBER 30, 1994

APPENDICES [Contained in Volume II]

Tables

Executive Summary

Beginning in the late 1940s and continuing for more than two decades, DDT and PCB's were released into the Los Angeles County sewer system and were eventually discharged through outfall pipes into the Southern California Bight (SCB) off the coast of Los Angeles. These chemicals settled to the ocean bottom, persisting in the sediments there and entering the food chain. The chemicals affected several species of local wildlife and continue to afflict at least four species — bald eagles, peregrine falcons, kelp bass, and white croaker.

The National Oceanic and Atmospheric Administration (NOAA) retained Natural Resource Damage Assessment, Inc. (NRDA) to undertake for the Trustees a study to estimate prospective *interim lost use value* (ILUV), a monetary measure of the compensation due the public as a result of these injuries. For the purposes of estimating prospective ILUV, the relevant time period was specified by the Trustees as 1994 to 2044. Therefore, the estimate of damages provided by this study does not include value for injuries occurring before 1994 and assumes that the affected species will recover naturally by 2044.

Given the specified injuries to the four species of SCB wildlife and the natural recovery time frame, NRDA determined that the best-available method for estimating prospective ILUV was contingent valuation (CV), a survey-based, economic methodology commonly used to construct economic values for a wide array of tangible and intangible objects. The theoretical foundation for developing valuation measures using contingent valuation is the same as that underlying all economic valuation regardless of whether the valuation is based on market transactions or non-market valuation techniques. In a CV study, as in all forms of economic valuation, the analyst observes a choice made by an individual and, from knowledge about that choice, constructs an estimate of economic value.

NRDA designed and implemented a CV study following best-available practices for survey design and administration. The object of choice described in the survey was presented in a referendum format in which respondents were given the opportunity to vote *for* or *against* a government program financed by a one-time income tax surcharge on California households. The program would reduce the time period for recovery of the four affected species from fifty to five years, *i.e.*, recovery would occur by 1999.

The aggregate estimate of prospective ILUV obtained from the study is $575 million (with a standard error of $27 million). This estimate is obtained by multiplying a per household ILUV estimate of $55.61 by the 10.3 million California households in the population to which the CV survey was designed to be extrapolated. The statistical approach used to obtain the per household estimate of value is a non-parametric maximum likelihood procedure developed by Turnbull (1976) which yields a lower bound on the sample mean. The resulting estimate was adjusted for respondents who did not pay California taxes by treating the *for* votes of non-taxpaying respondents as votes *against* the program.

The CV study on which this estimate is based is the culmination of an extensive program of instrument development — including focus groups, cognitive interviews, small pretests, and pilot studies — conducted over the course of 32 months, beginning in August 1991 and ending in March 1994, when the survey instrument was finalized. The main survey was administered over the next five months by Westat, Inc., one of the country's most respected survey research firms. Westat completed 2,810 in-person interviews with a random sample of English-speaking California households, achieving a response rate of 72.6 percent.

The reliability of contingent valuation surveys may be gauged by the questionnaire development, the survey implementation, and the properties of the results. In the current study,

the questionnaire development and the administration of the main CV survey adhere to the best standards in the disciplines of environmental economics and survey research. Our survey methodology was evaluated by comparison with the recommendations of the NOAA Panel on Contingent Valuation. The final approach adheres to NOAA recommendations or demonstrates that the approach used in the main survey is superior.

In order to assess the reliability of the results, relationships between respondents' choices and the following five groups of variables were examined: (1) the cost of the program; (2) respondent economic characteristics; (3) respondent preference-related and demographic characteristics; (4) respondent evaluations of the injuries and of the program to mitigate those injuries; and (5) respondent interest in, use of, and proximity to the affected natural resources. For both the pairwise approach recommended by the NOAA Panel and a multivariate approach, the variables hypothesized to be positively or negatively associated with the probability that respondents voted *for* the accelerated recovery program were found to be consistent with prior expectations. All measures with clear, expected relationships to respondents' preferences were statistically significant determinants of their choices. The judgments drawn from the multivariate model about the importance of particular variables in explaining respondents' choices are generally unaffected by modifications to the format used to represent these variables. For those variables which showed some sensitivity, the final specification remained the preferred model.

Qualitative data from the survey provided evidence that respondents paid attention to the survey and took the choice opportunity seriously, that respondents' decisions reflected their perceptions of and preferences for the object of choice, and that their choices were not influenced by extraneous factors. In particular, responses to the open-ended questions which asked respondents why they made specific choices suggest a good understanding of what the program

would accomplish and what it would cost.

The final basis for judging the reliability of this CV study relies on a demonstration that respondents are sensitive to the scope of the injury. This demonstration of sensitivity to scope is a key requirement of the NOAA Panel on Contingent Valuation for determining whether the results of a CV study are reliable enough for use in the estimation of natural resource damages. A second survey instrument (referred to as the *scope* instrument) was designed and administered in parallel with the main survey instrument (referred to as the *base* instrument) used to estimate the damages presented above. The scope instrument was identical in most respects to the base survey instrument with the exception that the injuries to the two bird species were excluded and natural recovery was described as taking place over 15 years rather than 50. To permit complete comparability between the results, these two survey instruments were randomly assigned and administered to two independent samples of respondents. The choices made in the two samples and the values constructed from those choices are significantly different, with the estimated lower-bound mean substantially smaller for the reduced set of injuries.

§ 1 Introduction

§ 1.1 The Initial Charge to Natural Resource Damage Assessment, Inc.

Natural Resource Damage Assessment, Inc. (NRDA) was retained by the National Oceanic and Atmospheric Administration (NOAA, hereafter, the Trustees) on May 6, 1991, to assist the Trustees in the preparation of a natural resource damage assessment.[1] The Trustees provided to NRDA a set of injuries to resources within the South Coast, an area near Los Angeles, geographically defined as lying within and along the northern part of the Southern California Bight (SCB).[2] NRDA was charged by the Trustees to undertake a study that would yield an estimate of *prospective* (*i.e.*, forward-looking) *interim lost use value* (ILUV) for losses due to these injuries to natural resources caused by DDT and PCB's released into the South Coast, where *interim lost use value* is defined as the amount of money required to compensate the public for injuries to natural resources.

---

[1] Other trustees involved in this assessment of damages are three State of California agencies and the U.S. Department of the Interior.

[2] The Southern California Bight is usually defined as that body of water lying within the curve of the California coastline between Point Conception and the California-Mexico border. For a more general description of the Southern California Bight and its resources, see Dailey, Reish, and Anderson (1993).

The time period of the injuries relevant for the estimation of ILUV is the time between the onset of injuries resulting from the discharge or release of the hazardous substance(s) and the time the affected resources are fully restored to their baseline conditions. *Prospective* interim lost use implies that the time period does not encompass injuries that have occurred prior to a specified date. In the context of this study, the date specified by the Trustees was March-August, 1994.[3] Prospective ILUV was to be estimated with reference to a specific natural recovery scenario, provided to NRDA by the Trustees, in which all injured resources would be returned to their baseline conditions over a 50 year time period.[4]

NRDA was instructed to assure that the method selected for estimation of prospective ILUV was consistent with the court opinion in *Ohio v. the United States Department of the Interior*[5] (hereafter, *Ohio*) overturning the Department of the Interior's (DOI) hierarchy of use values:

> Option and existence values may represent "passive" use, but they nonetheless reflect utility derived by humans from a resource, and thus prima facie, ought to be included in a damage assessment. [p. 464]

The *Ohio* Court stated that passive use value is a proper component of a natural resource damage claim.[6] Restated in economic terms, the court's opinion is that all aspects of a natural resource

---

[3] The main study field period; see Chapter 7.

[4] As noted, this 50 year natural recovery period and the injuries during that time that were to be valued were provided by the Trustees. NRDA was advised by the Trustees that the actual recovery period might differ from that specified and that the injuries we were directed to value were not necessarily all the injuries that occurred or all the injuries for which the Trustees would seek compensation.

[5] *Ohio v. U.S. Department of Interior*, 880 F.2d 432 (D.C. Cir. 1989).

[6] The *Ohio* Court's term "passive use" plus what is known as "direct use" combine to form

that give rise to well-being (*i.e.,* utility) derived by individuals are proper components of a damage claim. Thus, the appropriate measure of damages is the loss in *total* value due to the natural resource injuries.

Lost total value has been termed by many as the sum of use and nonuse (or equivalently, passive use) value. Whether one adopts the heuristic thinking of total value as the sum of use and nonuse, the most important feature of total value—consistent with the *Ohio* Court's definition of *use* to include passive use—is that it reflects all sources of value attached by individuals to a natural resource. Thus, the *Ohio* Court's definition of use (as equivalent to the sum of direct use and passive use) is labeled *total value* or, its equivalent in our framework, *interim lost use value*.

§ 1.2 Estimation Approach

what is known as "total economic value". Total economic value forms the basis for the interim lost use value presented in this report.

NRDA's study was designed to estimate prospective interim lost use value, the amount of money required to compensate the public for losses due to natural resource injuries resulting from DDT and PCB contamination in the SCB between March-August, 1994 and 2044, the time at which the resources were specified to be fully restored to their baseline conditions. Based on a consistent, welfare-theoretic definition for compensation,[7] NRDA estimated a lower-bound, monetary measure of required compensation. This measure is the aggregate of prospective total values lost by eligible California households as a result of the injuries.

§ 1.3 Contingent Valuation Method

NRDA determined that the best-available method for estimating total value, given the specified injuries and natural recovery time frame, was *contingent valuation* (CV). Contingent valuation is a survey-based, economic methodology that can be used to construct economic values for a wide array of tangible and intangible objects. CV is most often applied to the economic valuation of *public goods*, such as the value of improved air and water quality, and has been used for this purpose by numerous state and federal government agencies.

---

[7] See Chapter 3 for a discussion of this definition.

The concept of contingent valuation was proposed by Cirancy-Wantrup (1947). The first reported application was Davis (1963), which valued recreation in Maine. Since 1963, the number of published contingent valuation studies has grown rapidly with applications not only to environmental goods but also to other types of public goods.[8] The Carson *et al.* (1994) bibliography lists over 1600 studies and papers on contingent valuation. A large part of the growth of contingent valuation can be attributed to the use of contingent valuation by government agencies such as the U.S. Forest Service and the U.S. Environmental Protection Agency and by international organizations such as the World Bank and the Inter-American Development Bank. Currently, applications can be found from over forty countries.

The theoretical foundation of CV is the very same foundation underlying *all* economic valuation regardless of whether the valuation is based on market transactions or non-market techniques (*e.g.,* the travel cost method used to value recreational activities).[9] In all forms of economic valuation, the analyst constructs an economic value from an observed choice and from knowledge of the circumstances of that choice. All other things being equal, greater knowledge of the choice improves the validity of the constructed value. CV gives an analyst control over the choice presented to the survey respondent and over the circumstances in which the choice is framed. Other valuation methods usually rely on recorded past choices which require that the analyst make assumptions about features of the choice beyond his or her knowledge and control.

The design and administration of the CV survey described in this report were guided by multiple considerations including the Arrow *et al.* (1993) *Report of the NOAA Panel on Contingent Valuation* (hereafter, NOAA Panel), experience with past natural resource damage

---

[8] See Portney (1994) for a brief review of the history of contingent valuation.

[9] See Chapter 3 for further discussion of CV's theoretical foundation.

assessments,[10] experience in public policy evaluations involving non-market public goods, and

other research conducted by the principal investigators.

As part of the development of rules for natural resource damage assessment under the Oil

Pollution Act, Thomas Campbell, the NOAA General Counsel at the time this study was initiated,

formed a panel of social scientists to evaluate the reliability of CV for measuring passive use

values. The NOAA Panel was co-chaired by Nobel Prize winning economists Kenneth Arrow and

Robert Solow. Other members of the panel included three prominent economists—Edward

Leamer of the University of California, Los Angeles, Paul Portney of Resources for the Future,

and Roy Radner of Bell Laboratories—and the Director of the Survey Research Center at the

University of Michigan, Howard Schuman.

The NOAA Panel concluded:

> In Section IV above, we identify a number of stringent guidelines for the conduct
> of CV studies. These require that respondents be carefully informed about the
> particular environmental damage to be valued, and about the full extent of
> substitutes and undamaged alternatives available. In willingness to pay
> scenarios, the payment vehicle must be presented fully and clearly, with the
> relevant budget constraint emphasized. The payment scenario should be
> convincingly described, preferably in a referendum context, because most
> respondents will have had experience with referendum ballots with less-than-
> perfect background information. Where choices in formulating the CV
> instrument can be made, we urge they lean in the conservative direction, as a
> partial or total offset to the likely tendency to exaggerate willingness to pay.
>
> The Panel concludes 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. [Arrow *et
> al.*, 1993; p. 4610]

Many of the NOAA Panel's guidelines restate best-available practices pertaining to CV survey

---

[10] For example, see Carson *et al*. (1992).

design and administration, while a few are novel. In Chapter 4 of this report, we detail these guidelines; and in Chapters 5 and 7, we discuss the implementation of certain of these guidelines along with the other best-available practices implemented in this study.

This study also relies on additional information developed to gauge the soundness of the estimates derived from the CV survey, including (1) tests of various statistical hypotheses, (2) qualitative analysis of the data (*e.g.,* interviewers' evaluations of the survey's administration), and (3) the results of a split-sample test designed to evaluate the sensitivity of respondents' choices to the "scope of the environmental insult" (as suggested by the NOAA Panel).

§ 1.4 Peer Review

Richard C. Bishop, Trudy A. Cameron, and Alan Randall served as the primary peer reviewers for the overall study. Norman Bradburn, Norbert Schwarz, and Edward Tufte served in other various advisory capacities.

§ 1.5 Organization of Report

Chapter 2 describes the releases of DDT and PCB's into the marine environment of the Southern California Bight, the injuries caused by those releases, and the natural recovery process (as supplied to NRDA by Trustee representatives). Chapter 3 presents an overview of the economic concepts underlying monetary measures of value and describes the conceptual framework upon which this study is based. Chapter 4 describes the contingent valuation approach used for estimating interim lost use value and, using the NOAA Panel report as a template, addresses issues concerning the design and implementation of CV surveys and the reliability of

their results.[11]   Chapter 5 outlines the design and development of the two survey instruments (referred to as *base* and *scope*[12]) used in the main study, beginning with focus groups and ending with the fielding of the main study survey. Chapter 6 describes section-by-section the wording, format, and presentation in the base and scope questionnaires. Chapter 7 discusses the administration of the main study survey, including the sample design, interviewer training and supervision, quality control, completion rates, sample weights, and data entry. Chapter 8 discusses the qualitative survey data including responses to questions pertaining to respondents' choices and respondents' perceptions of the program described to mitigate the injuries as well as to interviewer debriefing questions. Chapter 9 discusses the statistical framework for the analysis

---

[11] Some confusion exists over the term *reliability* as applied to the results of a CV survey. As used by the *Ohio* Court and in the NOAA Panel report, the reliability of a measure is the degree to which it measures the theoretical construct under investigation. However, in the empirical social sciences, this preceding definition pertains to *validity*, whereas reliability is defined as the extent to which the variance of the measure is not due to random sources and systematic sources of error. While we believe that it is important to keep the distinction between validity and reliability clear, to avoid confusing readers who are unfamiliar with the social sciences, we follow the usage of the *Ohio* Court and the NOAA Panel and use the term *reliability* to refer to the degree to which CV surveys measure the theoretical construct under investigation.

[12] The *base* survey instrument was used as the basis for our estimate of prospective ILUV, while the *scope* survey instrument was used (along with the base version) in a test for sensitivity to the scope of the injuries. See Chapters 5, 6, 9, and 10.

and, within that framework, presents the quantitative results of the analysis. Chapter 10 presents the results of a test for sensitivity to the scope of the injuries. Finally, Chapter 11 presents the aggregate estimate of prospective interim lost use value.

§ 1.6 Title of Report and Study

The title of this report is *Prospective Interim Lost Use Value Due to DDT and PCB Contamination in the Southern California Bight*. Hereafter, for brevity, this report will be referred to as the *Lost Use Value Report* and the study on which it is based as the *Lost Use Value Study*.

§ 2 Injuries

§ 2.1 Introduction

The injuries valued in the Lost Use Value Study were provided to NRDA by Trustee representatives.[13] Those injuries related to the impacts of DDT and PCB's on the South Coast populations of Bald Eagles, Peregrine Falcons, White Croaker, and Kelp Bass. The description of the injuries and the context in which the injuries were placed were supplied by Trustee representatives or taken from commonly available public sources and verified by the Trustees.[14]

The Trustees stated at the outset of the study that modifications of the injuries to be valued should be expected; and, consequently, the study was designed to accommodate such changes. Because of these modifications, the description of injuries presented in the main study *base* questionnaire differed from those in some of the earlier, developmental versions.

The description of injuries appearing in the main study questionnaire was reviewed and approved by the Trustees prior to beginning the main study field period. The remainder of this chapter summarizes the injury description, including the context in which the injuries were placed, provides relevant excerpts from the main study questionnaire, and details the time frame for the

---

[13] As noted in section 1.1, the injuries presented to respondents in the main study base questionnaire are not necessarily exhaustive of either the set of all injuries caused by these releases of DDT and PCB's or the set of injuries for which the Trustees will seek compensation.

[14] The exact wording used to convey this information to survey respondents is presented in Chapter 6 and in Appendix A.1.

valuation.   Unless noted otherwise, the discussion pertains only to the injuries as they were described in the main study base questionnaire.

§ 2.2 Characterization of Injuries

The injury description did include all technical information available or conveyed to NRDA by the Trustees. As a practical matter, it is not possible to convey to respondents all the technical details of any given injury. Nor is it necessary to provide such details. What needs to be conveyed to a given respondent are the aspects of the injury that are relevant for that respondent's choice. Thus, one of the goals in designing the questionnaire was to describe to each respondent as completely as possible the information relevant to his or her choice.[15]

In addition to concerns about the relevance of information provided to the respondent, there is also a need for simplicity and brevity in conveying the injury to respondents. The relevant content and appropriate level of detail in the injury description of the main study questionnaire were re-evaluated throughout the development of the questionnaire.   When alternative characterizations of the injuries were available, the characterization which provided the more conservative[16] view of the injury was selected.

§ 2.3 Releases into the Southern California Bight

---

[15] See Chapter 5 for further discussion of this issue.

[16] That is, the characterization that would tend to cause no effect or that would tend to cause a respondent to vote *against* the offered program.

The description in the questionnaire of the releases of DDT and PCB's began with a statement that these two chemicals are found in the sediments on the bottom of the ocean off the Palos Verdes Peninsula.[17] The DDT came to be located in the sediments as a result of DDT manufacturing activities beginning in the late 1940's when a factory manufacturing DDT discharged waste DDT into the Los Angeles County sewer system. This waste DDT passed through a sewage treatment facility and was eventually discharged into the ocean through the treatment facility's outfall pipes. PCB's released by other sources also entered the marine environment through the sewer system. In the 1970's, sending DDT and PCB's into the ocean through the sewer system was stopped.

---

[17] A map showing the location and size of the deposit was shown to respondents. The main study base map and card set can be found in Appendix A.1.

§ 2.4 Contact with Natural Resources Other than Sediments

Small animals that live in the sediments absorb the DDT and PCB's as they feed in the sediment layer. When these animals are eaten by larger animals, the DDT and PCB's become bio-available to a wider group of animals, including two species of fish, White Croaker and Kelp Bass, and two species of birds, Bald Eagles and Peregrine Falcons. When enough DDT and PCB's accumulate in the bodies of the White Croaker, Kelp Bass, Bald Eagles and Peregrine Falcons, the chemicals impair the ability of these four species to reproduce. In the area of the deposit of DDT and PCB's, White Croaker and Kelp Bass produce fewer young than elsewhere along the California coast. In the 1950's, eagles and falcons in this area had trouble producing young primarily due to thin egg shells and, consequently, populations of these birds in the South Coast area disappeared. When adult falcons and eagles have been brought to the area from outside under controlled conditions, generally the newly introduced birds have not been able to hatch their eggs. Scientists believe that these reproduction problems are caused by the deposit of DDT and PCB's in the sediments off the Palos Verdes Peninsula.

§ 2.5 Injuries Presented in Main Study Questionnaire

As indicated above, the description of injuries in the main study questionnaire focused on injuries to the South Coast populations of four species: the Bald Eagle (*Haliaeetus leucocephalus*), Peregrine Falcon (*Falco peregrinus anatum*), White Croaker (*Genyonemus lineatus*), and Kelp Bass (*Paralabrax clathratus*). It also provided a description of potential substitutes for the injured resources. The following summary provides verbatim the excerpts pertaining to the injuries from the main study questionnaire:[18]

---

[18] See pages 5-17 of the main study questionnaire in Appendix A.1 for the full presentation of

Many species of fish and birds live off the South Coast. Four of these species are having problems producing young.

Two species of fish are having problems producing young in one place off the South Coast. These are White Croaker and Kelp Bass.

Two of the many species of birds living along the South Coast also have reproduction problems. They are Bald Eagles and Peregrine Falcons.

Many scientists have studied why these four species of fish and birds are having reproduction problems along the South Coast but not elsewhere along the California coast. They agree that these reproduction problems are caused by a deposit of two chemicals that are trapped in the sediment on the bottom of the ocean. These chemicals are DDT and PCBs.

The . . . scientists I mentioned earlier have conducted studies of the effect of this deposit. They know that DDT and PCBs can build up in the bodies of some fish and birds when the food they eat has these chemicals in it. According to the scientists, the only animals that are affected by this deposit are the four species I told you about.

They have found that the amount of DDT and PCBs in these two types of fish is so small that people would have to eat fish from this one area . . . on a regular basis to be harmed.

Fifteen years ago, the deposit of DDT and PCBs was also causing reproduction problems in several other species that sometimes feed in the area. However, these other species gradually recovered and now reproduce normally. Their recovery over the past 15 years was the result of a natural process. This process gradually covers the contaminated sediment on the ocean bottom with new sediment that is uncontaminated by DDT and PCBs. The deeper the contaminated sediment is buried, the more these chemicals are removed from the food these species eat.

Although these chemicals now no longer affect other species, they continue to affect the four species I told you about. Once the chemicals are buried deeper under clean sediment, these four species will also recover.

Until recently, there was no way to speed up this natural process. However, a procedure has now been developed to cover chemical deposits like this. If the State does not implement this program, nature will do the same thing, but it will take longer, 50 years instead of 5. That is, an additional 45 years.

---

the injuries.

§ 2.6 Time Frame for Valuation of Injuries

As indicated above, the questionnaire presented two alternatives for the recovery of the injured resources. The first alternative was natural recovery, which would take place over the next fifty years.[19] The second alternative, a recovery accelerated by covering up the described chemical deposit, would lead to recovery to baseline conditions in five years from "now" or 45 years earlier than that with natural recovery.[20] The respondent was presented with a choice between these two alternatives — the respondent was asked if he or she would vote *for* or *against* a program that would reduce the level of injuries occurring during the 50 years of natural recovery to the lesser level occurring in the five years of accelerated recovery. (Hereafter, this program is referred to as the accelerated recovery program.)

When based on this choice, the estimates of interim lost use value are prospective and do not include values for any of the injuries occurring before March 1994, *i.e.*, the injuries that predate the program, since the program would not prevent those injuries. Furthermore, estimates of prospective interim lost use value do not include values for the injuries which will occur in the five-year period of accelerated recovery despite the program. By definition, there will be some injuries until the end of the first five years even with the accelerated recovery program. Since these injuries occur with or without the program, value for them is not included in estimates of

---

[19] As noted earlier, this recovery time was provided by Trustee representatives.

[20] The Trustees have not indicated that they will actually implement such a program; the purpose of this program was to provide a plausible means of reducing the recovery time from 50 years to five years and thereby reducing the injuries.

prospective ILUV provided by the Lost Use Value Study. However, since the accelerated recovery program will prevent some of the injuries in the first five years as compared to natural recovery, the value of those prevented injuries are included in the estimates.

The relationship between the two recovery alternatives is depicted graphically in Figure 1. For natural recovery, the relative size of the injuries is represented by the area $a + b_1 + b_2$. For accelerated recovery, the area $b_1 + b_2$ represents the relative size of the injuries prevented, and the area a represents the residual injuries which occur despite the accelerated recovery. As noted above, injuries before 1994 are not considered.

§ 3 Economic Theory of Interim Lost Use Value (ILUV)

§ 3.1 Introduction

This study measures *prospective* interim lost use value (ILUV), a monetary measure of the public's losses due to natural resource injuries occurring over a pre-defined period.[21]   This measure corresponds to the compensation the public would freely accept in return for permitting a loss in well-being due to injuries to natural resources.[22] In this chapter, we discuss the theoretical foundation of prospective ILUV—the economic concept of value—and how prospective ILUV should *ideally* be measured.

Almost 50 years ago, the modern economic theory of consumer behavior provided a definition for monetary measures of economic value.[23]   An important step in the theoretical

---

[21] When the specification of the time period for damage assessment excludes losses due to past injuries, we have referred to this portion of ILUV as *prospective* interim lost use value.

[22] This is supported by the view of the *Ohio* Court where, during its discussion of contingent valuation as part of a damage assessment (section XIII of the opinion), it stated: "The purpose of such an assessment is to ascertain the amount of compensation due the public for an injury to the public's natural resources . . ." (p. 481).

[23] See Samuelson (1938 & 1948), Little (1949), and Houthakker (1950) for an early

development that led to this definition was the recognition that an individual's choices in markets

<u>revealed</u> that individual's preferences. Today, this choice-based theory of preference is a part of

most graduate textbooks in microeconomics.[24]

The economic concept of value stems from individual choices that involve trade-offs—

something is foregone to obtain something in return. Defining this trade-off in a particular way

allows the analyst to construct a measure of economic value for the *object of choice*. Objects of

choice may be either quite general or very specific and include the usual array of tangible goods

and services that we associate with market transactions. It is important to recognize that objects

of choice do not have to occur in specific measurable quantities, as apples or loaves of bread do.

They can range from services, such as a plane trip to a particular city, to broadly described states

of the world, such as experiencing a chronic disease or enjoying a scenic vista. As the object of

choice becomes more *intangible*, it can become more difficult to define units in which to measure

its quantity. Fortunately, to construct measures of economic value from people's choices, the

analyst does not need to describe these choices using neatly divisible units. Rather, to construct

economic value, all that is needed is a clearly identified (*i.e.,* well-defined) object of choice and a

specified consequence that results from the choice. Further, because economic values are

constructed from individual choices, the circumstances describing the context in which these

---

discussion of choice and revealed preference. At about the same time, J.R. Hicks wrote several
papers defining the theory underlying monetary measures of economic value. See Hicks (1939 &
1943).

[24] For example, see Varian (1992).

choices are made will be important to value measures. Taken together, the *object of choice* and the *circumstances of choice* fully embody the *choice elements* from which the analyst constructs values.

To provide a foundation for this chapter's later discussion of the measurement of interim lost use values, the definition of economic value in relation to the trade-offs associated with a choice is first discussed. After discussing how the rights to objects of choice can influence these value measures, this chapter describes the properties of economic values and how the measurement of ILUV imposes specific requirements on the construction of those values.

§ 3.2 Economic Concept of Value

§ 3.2.1 Background and Definition

The term *value* is used variously in everyday language as well as in many academic disciplines. Used as a verb, value conveys judgments of importance; as a noun, values can mean standards for evaluating behavior or factors contributing to personal or social well-being. Moreover, in each of these possible uses of the term *value*, there is clearly discretion in what each person describes as his or her values. In economics, however, value has a specific technical meaning; to define value in economics there must be a choice where something is given up to obtain something else, *e.g.,* an individual gives up a dollar to get a lottery ticket. Economists assume that people are able to consider objects of choice and, given their individual preferences, order them. Modern treatments of micro-economic theory now begin with choices as the basis for describing people's preferences.[25]

---

[25] See Kreps (1990) for a discussion of how choice can be used as a primitive concept and the

relationship between this logic and the usual development of consumer demand theory.

When a choice implies a consideration of alternatives, it defines a trade-off. The underlying economic theory of choice suggests that what is selected *must* be at least as desirable (from the perspective of the individual making the choice) as the alternatives that were not selected. Hence, the choice implies that the chosen object is at least *as good* (or *as valuable*) as what is given up. Thus, to assign a monetary measure of value to an object of choice does not require that people have dollar values for every conceivable object of choice in their consciousness. Rather, when a choice is made, the alternative that is foregone defines a <u>lower</u> bound for the value of the object selected. This lower bound is expressed in whatever units correspond to the alternative foregone. For example, if an individual chooses to give up a certain sum of money to obtain an object, that monetary payment represents a lower bound on the object's value to that individual given the circumstances of the choice. If the individual chooses to forego leisure time in order to obtain the object, its value would be denominated in leisure time. If the analyst wishes to determine how much the foregone leisure time is worth in money, (*i.e.*, to monetize the leisure foregone), additional assumptions would be required.

§ 3.2.2 Constructing Monetary Measures for Economic Trade-offs

To monetize economic value, the foregone alternative (defined by an individual's choice within a specified trade-off) must be expressed in dollars. Unfortunately, this monetization has sometimes created misconceptions. For example, it has been suggested that economic values are confined to prices observed in markets.   These misconceptions arise because many people commonly think of the monetary measure of economic value as a *price*—if a widget sells for $6 in a market, then $6 must be its value. This view is misleading, however. When a person buys a widget the analyst only learns that it is worth <u>at least</u> $6 to the buyer. He or she might be willing to pay much more than $6 if necessary to get the widget. Markets do offer opportunities for people to make choices but, it is these choices and the circumstances relevant to them, that permit construction of the underlying economic values, not the market prices *per se*.

Any time a person makes a choice and a trade-off is defined, an economic value may be constructed. Of course, the existence of trade-offs does not guarantee that the analyst is aware of all of the elements of the choice, including the consequences of the decision and the alternatives foregone in favor of an individual's observed choice. Ideally, to construct a value, the analyst would have complete knowledge of the choice elements. When information is incomplete, as is usually the case with indirect methods for valuing natural resources (discussed below), the analyst must supplement what is known about the choice elements with assumptions.   Formal definitions of a monetary measure of economic value require a specified assignment of rights (*i.e.,* some degree of control over the object of choice resides with a specified individual or group of individuals).   The assignment of rights is essential to a choice because each choice involves receiving something and, in return, giving something up. Thus, we can define this implicit trade-off in two ways: (1) giving something up to receive the chosen object, which corresponds to the

willingness-to-pay (WTP) concept, or (2) receiving something to give up the object, which corresponds to the willingness-to-accept (WTA) concept. The first of these measures, WTP, is relevant when the individual does not have rights to the object of choice, while WTA is relevant when he or she does.

WTP and WTA are the fundamental monetary measures of value in economics. *All* economic valuation can be shown to correspond to one or the other. In a situation involving a WTP choice, the economic value of the object of choice is constrained by the *wealth* of the individual (*i.e.,* the personal possessions that an individual can give up). Thus, while an individual may in theory be willing to give up all of his or her wealth to obtain some highly desired object of choice, the upper bound on an individual's economic value for the object is constrained by how much wealth each person has and is able to give up.

Constraining the economic value of an object of choice to available wealth does not arise in a situation involving a WTA choice. Here the item to be valued (or object of choice) is something the individual already possesses (*i.e.,* the relevant right resides with that person) and the item is a part of his or her possessions. The trade-off requires the specification of *something* the individual will freely accept in exchange for the object of choice. Because, in this situation, the object of choice is already part of the individual's possessions, the *something* an individual will freely accept in return for that object is not constrained by that individual's wealth. Thus, the monetary value of an object of choice constructed from either a WTP or WTA choice can differ.

To describe more formally the connection between trade-offs and monetary measures of WTP and WTA, consider a simple choice relevant to each. In the case of WTP, the choice is the opportunity to acquire something. Objects of choice can be very general. For example, the object offered could be an improvement in air quality or it could be a different pattern of community land

use. In this latter example, WTP is the maximum amount of money each individual would be willing to give up to avoid having a particular pattern of land development. The development may be on land the individual does not own, but nonetheless influences his or her activities or well-being. If we observe choices where the individual can attain the object by foregoing less than the maximum amount that he or she is willing to pay, the amount foregone is a *lower* bound on WTP.

WTA involves a different type of choice. Here we consider a person agreeing to give up the object of choice in exchange for an increase in monetary wealth. Selling anything places people in this choice situation. In the land development example, one way a WTA choice would arise is if the individual owned the parcel required for the development. Since the individual requires something to voluntarily agree to give up the parcel of land, the measured economic value of the parcel could be different in this WTA setting.

Examples are of course not limited to land. When people accept a job (*i.e.,* sell their labor services), sell a house or a car, or agree to permit a neighbor to modify his or her house so it would then block part of a scenic view, the choices involve a WTA trade-off and, hence, allow the construction of an economic value for the object of choice given up. If the choice posed to the individual involves a monetary payment in exchange for the object of choice, then the compensation payment provides the basis for constructing the monetary value of the object of choice. If the payment to the individual is greater than the minimum that the individual would be willing to accept, then the payment is an *upper* bound on WTA. If the payment should happen to be the minimum amount the individual would accept to give up the object, then that amount is WTA.

The assignment of rights specified by the circumstances of the choice determine which

measure (WTP or WTA) is the appropriate basis for specifying the monetary measures of economic value.[26] To proceed from these definitions to actual measurement approaches requires further assumptions to which we now turn.

§ 3.2.3 Economic Values

In order to construct a monetary measure of the current value of an object (given either a WTP or WTA trade-off), the analyst must: observe a choice relevant to the object of interest; understand the circumstances of that choice; and link the choice outcome to monetary implications. The *circumstances of choice* include the assignment of rights, the alternatives foregone/accepted, the choice mechanism, the certainty associated with the choice outcome, and other features of the choice relevant to the individual's decision-making.

The assignment of rights, discussed above, refers to the degree of control the individual has over the object of choice. In the case of very simple private goods, such as a painting, the individual either owns the painting (*i.e.,* possesses all relevant rights to the painting) or does not own the painting (*i.e.,* possesses no rights). In the case of some public goods, such as local parks, a person may possess a right to use the park, but each individual is not free to sell the park.

The list of alternatives foregone/accepted pertains to both WTP and WTA choices. In the case of a WTP choice, the alternatives foregone represent those objects in the possession

---

[26] It is generally accepted that for frequently traded market goods subject to price changes, differences between WTP and WTA may be small (see Willig, 1976). However, for quantity changes in public goods, the differences between WTP and WTA can be quite large (see Hanemann, 1991).

of the individual that are given up to obtain the object of choice. On the other hand, in a WTA choice, the alternatives accepted represent those objects not in the possession of the individual that are accepted in return for the object of choice.

The choice mechanism is the institution or set of rules that provides the structure for the execution of the choice. For frequently traded private goods (*e.g.*, milk), the choice mechanism is often an established market, while for infrequently traded goods with few buyers and sellers (*e.g.*, large construction projects), the choice mechanism may be a negotiated contract between a buyer and a seller. In the case of public goods, such as increased fire protection, the choice mechanism may be a local referendum.

The certainty associated with the choice outcome refers to the perceived certainty from the individual's perspective about how likely it is that the object of choice will be obtained/given up and the alternatives foregone/accepted. For example, when a person votes to have his or her property taxes raised to support improved fire protection, he or she has some idea in mind about how likely it is that he or she will receive improved fire protection.[27] The degree of certainty may affect each person's choice and thus the value an analyst can construct based on that choice.

Finally, other features of the choice relevant to the individual's decision-making are usually a composite of other factors specific to each decision. One important factor in this category (discussed below) concerns whether the choice in question is part of a sequence of choices and where in that sequence the choice in question is placed.

---

[27] Some individuals may also form perceptions about the precise amount of tax they will be required to pay and base their vote on their expected payment.

Absolute control by the analyst over the circumstances of choice would greatly facilitate this process of value construction when using indirect approaches[28] (sometimes referred to as revealed preference methods); however, analysts are unable to exercise such control. Revealed preference approaches, such as the travel cost method, use observed <u>past</u> choices and attempt to gather all relevant information *(i.e.,* elements of the choice) pertinent to those past choices. Because (a) all of the pertinent elements of past choices can *never* match exactly the choice involving the object we currently wish to value (*i.e.*, at the very least, time has elapsed since previous decisions), and (b) the records of past choices are often incomplete, the analysis usually rests on important assumptions that are introduced to make past choices relevant for the current valuation. For example, in the context of valuing natural resources with the travel cost method, an analyst could observe recreational uses of a particular beach and then make assumptions about all pertinent choice elements that lead an individual to use the beach for recreation. These assumptions would include such things as: what each individual perceived as his or her costs to use the beach;[29] how much each person used the beach; what they might have considered as

---

[28] The term *indirect* is used with revealed preference because the choices used to construct economic values for non-marketed resources are observed. The object of choice and circumstances of choice must be specified as assumptions by the analyst based on what is known about the types of resources involved and what can be observed.

[29] For example, the costs of a trip to the beach could include the vehicle operating costs and the time costs associated with the travel. The measurement of the time costs also requires

substitutes; and, perhaps most importantly, whether there would be reasons for individuals to be

concerned about the status of the beach beyond the observed pattern of past use.

---

assumptions about how the time would otherwise be used. If the person had the opportunity to work for pay, the appropriate cost might be the wage rate. If not, further assumptions would be required to define the relevant cost.

Concerns for the status of the beach beyond direct use give rise to what has been called passive use. Since individual concerns giving rise to passive use are not observable from behavior,[30] analysts choosing to employ revealed preference approaches assume by default that passive use contributes nothing to the total economic value. This follows because the choices recorded in a revealed preference format are confined to situations with direct use of the resource.

Thus, the revealed preference approach to valuing a natural resource has several important limitations. First, as suggested above, it can only use choices that can be observed and recorded; for instance, in the above example, the object of choice is limited to observed recreational use. Second, and equally important, the metric it uses to quantify the object of choice is restricted to what can be conveniently measured.[31] Continuing the travel cost example, the object of choice implied by a decision involving recreation at a beach is typically quantified by analysts using indirect methods such as the number of trips made to the beach during a recreation season. Finally, since the analyst's definition of the object is derived solely from observed past behavior, which can only reflect direct use, revealed preference approaches (by design) measure only a

---

[30] Some individuals never go to a particular beach, but are concerned about its status. Simply observing patterns of beach use alone would overlook these passive use concerns.

[31] This arbitrary unitization (*i.e.,* expressing the object of choice in some unit of measurement) leads to a corresponding arbitrary *value* unitization (*e.g.,* dollars per beach visit). Unitization is not required by the economic definition of value; rather, it is performed at the will of the analyst and may be irrelevant to the circumstances underlying the actual choice made by the individual.

portion of total economic value when passive use concerns are present; observable actions provide an incomplete picture in that they only measure a particular use value.

This conclusion follows from a closer examination of what is implied when recreation trips are used as the exclusive basis for describing an individual's interest in the beach. If the goal is to measure the economic value of the beach, then the analyst has imposed a potentially false equivalence — the observed trips are equated with the way each person is assumed to conceive of the beach as an object of choice. What is actually observed, decisions about recreation trips in a given time period, must be interpreted by the analyst as providing a complete record of the circumstances of choice for the beach. If the concern involves the beach as a natural asset, decisions to use it at a particular time offer only one type of choice involving the beach.

The nature of this distinction can be seen in the following example. Suppose that an oil spill sufficiently contaminates the beach so that it can no longer be used by the public. In this case, the ideal choice would involve a negotiated WTA compensation (discussed further below) for the temporal pattern of injuries due to the spill and each individual would evaluate the complete object of choice as it is known to him or her (*i.e.*, the temporal pattern of the injuries). In contrast, by using the individual's past recreation decisions as equivalent to this ideal choice, the analyst is assuming that the relevant object of choice for the injuries to the beach is confined to an observable measure, the trips to the beach that would be precluded by the temporal pattern of injuries. Using the revealed preference approach, economic values are constructed from choices comparing objects of choice with and without the opportunity to take these trips, not objects of choice with and without the injuries. Hence, observed behavior necessarily limits what can be inferred about people's values.

An alternative approach that does not rely on the objects of past choices as proxies for an

object to be valued at present is one that offers individuals an object constructed to be identical (or as identical as possible) to the object the analyst seeks to value. This approach is embodied in *contingent valuation* (CV). Two issues are important to this approach. The first is the degree to which the object offered corresponds with the object for which the analyst wishes to construct a value. Because this approach permits control by the analyst over the elements of choice, the degree of correspondence can be very close and the object and circumstances of the choice can be structured so that the choice provides information from which the *total* economic value of the offered object may be constructed.

The second is the degree to which individuals making the choice accept the financial responsibility for the consequences of these choices. Because individuals make the choice in the context of a survey, it is critical that individuals accept the responsibility for the consequences of their choice. That is, in a WTP context, they should accept the responsibility for making the financial payment should they wish to obtain the object offered. Or, in a WTA context, they should be willing to forego the object of choice should they receive a compensatory payment. To ensure such acceptance on the part of the respondent, the offered choice—including the assignment of rights, alternatives foregone, choice mechanism, and the certainty associated with the choice outcome—must be plausible.

§ 3.2.4 Reliability of Monetary Measures of Economic Value

Since all monetary measures of economic value are constructed by analysts from people's choices, it is natural to ask whether there are ways to evaluate the validity (reliability)[32] of what has been measured. Attempts to judge validity necessarily rely on indirect evaluations. In the case of revealed preference methods, an economic value derived from observations of people carrying out an action that is hypothesized to be related to a natural resource (*e.g.,* taking a trip to the beach) is assumed to be valid by economists because an action took place. That is, an individual chose to use the beach. Clearly, if the assumptions linking the observed behavior to the object of choice (*e.g.,* recreational beach use) are correct, the action signals an individual's interest in the beach. But, one should not assume the economic value constructed from this choice represents a total value. People taking the observed actions may not necessarily interpret the elements of choice in the way the analyst has assumed. Nor is it guaranteed that assumptions made by the analyst regarding the elements of the choice are true. At best, the object of choice that can be valued from such observations is the specific action observed. Moreover, the fact that a choice was actually made by someone does not validate a measure of economic value derived from it. It simply reflects the relationship between the actual choice and the constructed economic value.[33]

Like revealed preference approaches, external validation of CV estimates of value must also rely on indirect evaluations. The validity (reliability) of CV estimates of interim lost use value is addressed in the next chapter.

---

[32] See Chapter 1, footnote 11.

[33] For an overview of these issues in the context of a travel cost model, see Bockstael, McConnell, and Strand (1991) and for a critique of the assumptions used in such models, see Randall (1994).

§ 3.3 Interim Lost Use Value, Negotiation, and Damage Assessment

Constructing a valid monetary measure of the public's losses due to the injuries to natural resources resulting from releases of DDT and PCB's requires that these injuries be conveyed in a *credible* choice context with *recognized* consequences.   The appropriate choice is one that permits the construction of a *total* economic value. An *ideal* choice would be one in which each member of the public (acting through agencies that serve as trustees) would agree to "permit" a pattern of injuries, restoration, and recovery. This choice defines the trade-off desired by isolating what must be given to each person (monetary compensation) for him or her to freely forego the object of choice (in this case, the resources in their baseline states). Each person's acceptance of compensation implies an upper bound on economic value for the object of choice.

This description of ILUV is analogous to what would be sought if the trustees for the natural resources and those responsible for the injuries could *negotiate* a payment in advance to assure that people's losses would be compensated. Such a negotiation might be similar to actual negotiations one might observe between local political jurisdictions and private parties seeking to site unwanted facilities (*e.g.*, land fills) within the jurisdiction. In these idealized negotiations, public officials, acting on behalf of the public, and the private party wishing to locate a facility negotiate a monetary payment sufficient to compensate the public for disamenities that will be brought about if the proposed facility is sited.

It is difficult within the framework of a CV study to construct a WTA choice that directly parallels the choice and trade-offs described by the negotiation example. The elements of such a choice would require people to envision a well-defined set of injuries, restoration, and recovery *before* the injuries have taken place. This complete package would need to be offered as a choice in advance of the actual injuries to capture fully the intention of the negotiation metaphor. To

implement this WTA choice in a CV survey for a natural resource injury that has already occurred, it would be necessary to ask respondents to imagine a situation where they were offered a choice that would allow them to prevent the <u>actual</u> injury <u>before</u> it happened. In essence, one would be designing a survey instrument that asks the respondent to mentally travel back in time to a point just prior to the event that caused the injuries and pose to the respondent a WTA choice.

§ 3.3.1 Defining Credible Economic Choices

In order to ensure that respondents take the choice in a CV survey seriously, it is highly desirable that the mechanism by which the object of choice would be provided and the payment obligation be plausible. In practice, it is very difficult to design a survey that makes a "time traveler" choice (described above) credible to the respondent.[34] For this reason, CV surveys have tended to use two alternative approaches to the construction of measures of economic value for natural resource injuries. In the first approach, the object of choice is presented as a program to prevent *future* injuries; these injuries are described identically to the actual injuries in question. For example, the economic value for injuries caused by an oil spill might be measured from the choices people make when those injuries are presented as the anticipated result of not implementing a prevention program. In the second approach (and, as Chapter 4 describes, the form adopted for this study), the object of choice is portrayed as a program to <u>alter the recovery</u> of the injured resources. In this instance, the program enables the resources to return to their baseline (*i.e.,* original) conditions more quickly than they would if nature took its course.

---

[34] This is why the NOAA Panel advises against using the "conceptually correct" measure "of the minimum amount of compensation that each affected individual would be willing to accept" (Arrow *et al*., 1993; p. 4603). The Panel also noted that the WTP approach (used in this study) is "the conservative choice" (Arrow, p. 4608).

The requirement that a CV survey offer respondents a credible choice has two key implications. First, it shifts the focus of the object of choice from the existing resource injuries to the ways a program might address those injuries (*i.e.,* how the resources might be returned to their baseline states). Second, to define a trade-off, the choices involving the program must have consequences for the people asked to make a decision about the program. Imposing consequences for respondents that result from their choices (*i.e.,* requiring them to forego something in return for the program) changes the perspective from compensation, or WTA, to WTP.

Both of these implications—the shift in focus and in valuation perspective—modify the elements of the choice and thereby affect the economic value measure in such a manner that it would yield a lower bound on prospective ILUV, not an upper bound. The first implication imposes limits on the program designed to influence a set of injuries (*i.e.,* the program must be credible), and the second alters the presumed rights for the resources. The injuries have already taken place and are continuing to occur. Hence, the individual cannot usually be placed in a situation where the *anticipated* losses from those injuries can be offered to him or her as a plausible choice (as in the "time traveler" choice referred to earlier). This implies that the object of choice offered is the alteration of the time path to recovery. Control of that outcome requires a payment and, thus, a WTP perspective. Coupled with this perspective is the income constraint (here used synonymously with wealth) — a constraint on how much of an individual's income can be part of the trade-off defined by each person's choice at the time that choice is offered. While these adjustments are compromises serving to introduce limitations in the monetary measure of economic value, the direction of their impacts is consistent and known — they generally serve to understate the monetary measure of economic value.

The ideal choice, described by the trustee–responsible party negotiation prior to the injury, has implications for how a WTP choice provided in a CV survey should be implemented. Three issues are especially important: (1) the role of the information provided to CV survey respondents, (2) the importance of sequencing objects of choice as part of measuring monetary values for any particular object, and (3) the role of nesting as a logical issue in defining the object of choice.[35] Each of these is discussed below.

---

[35] The term "embedding" has often been used to refer to a variety of unrelated phenomenon in the literature. It sometimes is used to refer to a relationship between one object of choice and other objects that could be complements or substitutes. In this usage, the discussion often deals with the impact of changes in sequence in these objects and its effect on constructed economic values. A second usage relates to natural groupings or aggregations of particular objects of choice with appropriate subsets nested in more encompassing categories. In what follows, the first usage will be addressed as issues in sequencing and the second, as nesting. For a discussion of these issues, see Carson and Mitchell (forthcoming).

§ 3.3.1.1 The Role of Information

It has been suggested that since respondents to contingent valuation surveys would not know about the injuries to the resource of interest if not given information about them during the interview, providing that information induces value in people who do not value the resources. Under this view, only those individuals who know of the resource in its baseline state and of the injuries *prior* to a contingent valuation interview could experience a loss in well-being and be considered in a total value ILUV calculation. This argument is incorrect. For example, when one considers how negotiation prior to an event that causes injuries would take place, it is natural to conclude that any negotiated compensation would require information be provided to the public prior to a decision. The anticipated pattern of injuries must be known for individuals to freely agree to permit the losses arising from injuries to natural resources (*i.e.,* the interim lost use). However, the public would not necessarily be expected to know any relevant aspects of the object of choice prior to the time they are offered a choice as part of the negotiation. It is reasonable to conclude that the issue of prior information would extend to all negotiations. For example, the public may not be aware of all disamenities they would experience if a landfill were sited in their locale; however, before agreeing to a compensation package they would want to be informed. The same logic holds for damage assessment.

When the object of choice is a program to affect the recovery of the injured resources and the terms of the economic trade-off involve a WTP rather than WTA perspective, the information requirements are not altered. Relying on this type of WTP choice does not dictate a change in, or allow the analyst to ignore, the elements upon which the ideal choice is based. Rather, the choice used for value estimation should mimic as closely as possible the ideal choice and this requires that the public be informed.

§ 3.3.1.2 Sequencing

On occasion, it has been argued that the choice posed in a CV survey designed to estimate the total value of injuries to natural resources should be placed in a sequence of other choices.[36] That is, the choice designed to value the injuries of interest would follow a choice designed to value another object (*e.g.,* the provision of homeless shelters). Since the order in which an object is valued is one circumstance of the choice, the value constructed from a particular choice will depend on that choice's order in the sequence (*e.g.,* if a choice is offered later in a sequence, this generally increases WTA and lowers WTP).

As noted above, the ideal choice envisions the public receiving negotiated compensation for losses experienced until the affected resource is fully restored. Elements of this ideal choice do not entail asking the public to select among alternative public projects (*e.g.*, homeless shelters); nor do they require them to participate in the allocation of public funds among other private or public goods.

While sequencing effects might be important conceptual issues in developing monetary measures of the economic value of investments to preserve <u>new</u> natural areas, they are irrelevant to damage assessment. For instance, consider a market example where someone approaches you to buy your car but, before you negotiate the price of the car, the party wants to negotiate a price for your house which you don't want to sell. It is hard to see how such negotiations would develop a more reliable measure of the economic value of the car given the very different elements of choice.

---

[36] For example, see Kemp and Maxwell (1992).

§ 3.3.1.3 Sequencing with Nested Objects of Choice

Sequenced decisions involving nested objects of choice, assumed to be larger or smaller along some scale, change the circumstances of choice and therefore the constructed values. For example, if natural resource injuries in question pertain to a single oil spill, that object of choice could be nested in injuries from the set of all oil spills. This set would contain the spill in question along with others. Sequenced decisions with nested objects of choice are not relevant for the assessment of interim lost use. In a negotiation framework, where compensation for injuries is negotiated <u>prior</u> to the injuries occurring, decisions involving larger injuries are simply irrelevant.[37]

---

[37] Nesting objects of choice is relevant to testing the sensitivity of respondents to the scope of the injury. See Chapter 4 for a discussion of the NOAA Panel guideline on testing the sensitivity to scope.

§ 3.4 Summary

The monetary measure of total economic value introduced here and applied in subsequent chapters to the quantification of ILUV corresponds to the conventional measure of economic value defined over fifty years ago. At that same time, the link between people's preferences and the choices they made was explicitly recognized. Adaptations of that theory to situations involving non-market valuation started with choices giving rise to *observable* actions. However, these observable actions often provide an incomplete picture of how natural resources enhance people's well being. This is because they focus only on a subset of the people (users) who might care about the resources in ways that lead to observable actions and don't capture any other reasons why people might be concerned about the resources.

The logic underlying the measurement of *ILUV* requires a method that can capture all of the reasons why people would be concerned about injuries to the affected resources. In the ideal situation, they would be presented with a choice involving a specific pattern of future injuries, natural recovery from those injuries, and a possible program of restoration. Their choices would provide information that describes the losses people would experience as a result of these injuries and would necessarily be part of a negotiation between the trustees for the resource and the private party who desires to "use" the resource in a way that would lead to injuries. This negotiation logic is consistent with the requirement that ILUV be based on compensation and implies that WTA is the appropriate valuation perspective.

Frequently, this ideal WTA choice framework cannot be implemented in practice. Nonetheless, the choice used for value estimation should mimic as closely as possible the compensation required by the public to accept the temporal pattern of injuries. The negotiation

perspective clarifies the degree to which respondents should be informed. It would be inappropriate to expand the set of injuries to include the possibility of other losses not actually being offered, thereby nesting the injuries actually under evaluation in a larger context, because the larger set of injuries would be outside the domain of relevant negotiation. Likewise, the negotiation terms would not be contingent upon decisions made by other parties to enhance or reduce the other types of resources available to people. Simply stated, the decision, out of necessity, to adopt a WTP perspective for damage assessment, does not imply that one must adopt all the trappings of a WTP framework suited to the valuation of new public goods.

§ 4 Measurement of Interim Lost Use Value (ILUV)

§ 4.1 Introduction

Building on the theoretical framework presented in Chapter 3, this chapter discusses in more detail the approach chosen to measure prospective interim lost use value (ILUV) — the contingent valuation (CV) method. An overview of the format of the CV questionnaire used in the Lost Use Value Study is provided as a foundation for the following discussion on the issues concerning the reliability of CV estimates of ILUV.[38] The NOAA Panel report is used as an organizing template in this latter discussion. Finally, this chapter systematically addresses within the context of this study each of the NOAA Panel's specific recommendations.

§ 4.2 Choice of Approach

The task of quantifying the compensation required by the public for the losses due to injuries to natural resources was addressed at a conceptual level in section 3.3:

> Constructing a valid monetary measure of the public's losses due to the injuries to natural resources resulting from releases of DDT and PCB's requires that these injuries be conveyed in a *credible* choice context with *recognized* consequences. The appropriate choice is one that permits the construction of a *total* economic value. An *ideal* choice would be one in which each member of the public (acting through agencies that serve as trustees) would agree to "permit" a pattern of injuries, restoration, and recovery. This choice defines the trade-off desired by isolating what must be given to each person (monetary compensation) for him or her to freely forego the object of choice (in this case, the resources in their baseline states). Each person's acceptance of compensation implies an upper bound on economic value for the object of choice.

This framework implies that interim lost use value corresponds to the compensation required by the public for losses due to injuries to natural resources, and that this compensation can be

---

[38] See Chapter 1, footnote 11.

measured by the sum of the monetary payments individuals would require to freely accept the loss in well-being brought about by the injuries. In practice, to meet this conceptual objective, we must ask two questions: (1) can we, as analysts, observe an actual historical choice where the elements of that choice match the ideal choice described above? and, (2) can we construct a total value that corresponds to ILUV from those observations of individuals' choices?

Section 3.3 suggested that the ideal conceptual choice for damage assessment can be understood through a metaphor — how the trustees and the responsible party would negotiate a compensation payment in advance of the release giving rise to the injuries. While one might find in the historical record examples of such negotiations in other situations, to our knowledge, no such prior negotiation occurred with respect to the injuries of concern in this damage assessment. It was also noted in Chapter 3, that the task of constructing a credible WTA-choice, paralleling the negotiation metaphor, was deemed problematic and a choice based on WTP for an alteration in the time path to natural recovery was adopted instead. One could ask whether there exist historical choices corresponding to the WTP-choice adopted for the Lost Use Value Study. To the best of our knowledge, no such historical choices exist.[39]

As indicated earlier, the approach adopted for the measurement of required compensation

---

[39] As noted, numerous problems arise whenever one attempts to use revealed preference approaches to estimate total value. The most serious problem arises from the fact that the object of choice can only be defined by the historical choices that pertain to observable outcomes (*e.g.*, recreation), and the analyst must select one such outcome as the basis for the value construction. This necessarily results in an object of choice with attributes that imply the constructed value corresponds only to use value.

due to injuries to natural resources in the Southern California Bight was contingent valuation (CV). We now turn to a discussion of this methodology.

§ 4.3 Contingent Valuation Method

The contingent valuation 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.[40] Additional 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.

The CV method has increasingly become accepted for measuring the benefits of policy actions and thereby used to inform public policy decision making. In 1979, the Water Resources Council included CV as one of three recommended methods for determining the benefits of federal water and related land resource projects. Since that time, various federal and state agencies have used the method for policy purposes and, as mentioned, it has been recognized by the *Ohio* Court and the NOAA Panel (Arrow *et al.,* 1993) as a method capable of providing useful information for the evaluation of natural resource damages.

The contingent valuation method uses the same logic that underlies the definition of the monetary valuation concepts discussed in Chapter 3, that is, choice. In CV studies, choices are posed to people in surveys; analysts then use the responses to these choice questions to construct

---

[40] See Portney (1994) for a brief history of contingent valuation.

monetary measures of value. The specific mechanism used to elicit respondents' choices can take a variety of forms, including asking survey respondents whether they would purchase, vote, or pay for a program or some other well-defined object of choice. It can also be a direct elicitation of the amount each respondent would pay (WTP) to obtain an object of choice or the amount each respondent would accept in compensation (WTA) to give it up.

When used for damage assessment, a contingent valuation survey presents each individual with an opportunity to make a choice, where the object of choice is usually a plan or program to prevent the relevant injuries or to restore the affected resources to their baseline conditions and thereby mitigate the relevant injuries. The context for that choice can be any setting that is regarded as credible by survey respondents. Because the elements of the choice can be presented in some detail to each respondent, there is no need to rely on historical choices and impose the assumptions required to link those choices to the object to be valued.[41] Rather, in damage assessments relying on CV, respondents' choices are directly linked to the object of choice (*e.g.*, through a specified restoration or prevention program that addresses the relevant injuries), thereby enabling the analyst to construct the appropriate measure of *total* value.

In a CV survey, respondents are presented with material which can be described as three separate (but integral) components:

1.   The key elements of a CV survey are the object of choice and the circumstances of the choice (including the method proposed for each individual to pay or receive compensation, the time period over which the object or decision is relevant, the relationship of the object to available

---

[41] See Chapter 3 for further discussion of the nature of the assumptions required when the analyst utilizes indirect or revealed preferences approaches to construct values for objects of choice.

substitutes, and other elements of the choice relevant to decision-making). When using CV to construct ILUV, the object of choice can be a program or a set of activities to restore the injured natural resources to their baseline conditions. To understand the program, each respondent must understand the nature of the resources, the character of the injuries to those resources, and any natural recovery process that might influence how these injuries would be mediated in the absence of undertaking the program.

2.   After the description of the *object* and *circumstances of choice*, the CV survey elicits a <u>choice outcome</u> which can be used to construct each individual's total value for the object of choice that has been presented. The choice can be a direct elicitation for a single value or repeated questions for an interval estimate of the value. In all cases, it will describe a *specific choice mechanism* and elicit choice information used to construct value.

3.   Because individuals have different preferences and face different constraints, questions are also asked about respondents' attitudes, social characteristics (*e.g.,* age, education, gender, race, etc.) and economic characteristics (*e.g.,* income).

We noted in Chapter 3 that an ideal CV choice is one that is framed in a WTA context in which a respondent is offered the opportunity to consider a situation where the injuries had not yet occurred and to choose how much compensation would be required to permit the time path of recovery associated with the injuries. Because the injuries in the South Coast have <u>already</u> occurred over a period of years, the only way to implement this approach would be to ask the respondents to assume that they could mentally travel back in time to before the injuries occurred. Such an approach is difficult to convey in a way that induces respondents to accept financial responsibility for a program to respond to the injuries. Rather than pursue the WTA framework, we have relied upon monetary values constructed from a contingent valuation choice based on WTP for a program that would speed-up the natural restoration of the injured resources. For reasons explained in Chapter 3, a WTP choice provides a lower bound for values that would have been constructed from a WTA choice.

§ 4.4 Format of a Contingent Valuation Survey

In Chapters 5 and 6, we review the design and format of the main study survey, but a working knowledge of the survey's structure is useful as a basis for understanding the issues of CV reliability that we discuss later in this chapter. Thus, this section provides a brief introduction to the structure of a CV survey in general and the instrument used in the Lost Use Value Study in particular.

The first part of our CV survey introduces the general topic of the survey. The topic had been described to each potential respondent in an advance letter and by the interviewer in the course of seeking the respondent's participation.[42] Following this initial description, the CV survey introduces the specific topic of the survey and then provides the information describing the elements of the choice. From this point CV surveys and general public opinion surveys differ in fundamental ways.[43] CV surveys typically focus on a single situation which is described in some detail. In a damage assessment, the elements of choice that are presented include a carefully

---

[42] The advance letter stated: "Westat, Inc., a survey research firm, is helping the State of California conduct a study about the opinions of Californians on issues facing the state today such as education, the environment, and crime"; see Appendix C.2.2. The interviewer's pre-scripted introduction essentially repeated the information contained in the advance letter; see Appendix C.2.1.

[43] Typical of public opinion surveys are those conducted for newspapers which ask respondents a few questions about each of a number of current public issues and political figures.

worded description of the natural resource and the injuries. As noted above, this description presents the resource and injuries so that the respondent will perceive his or her relationship to available substitutes and other factors relevant to the individual's choice.

The description of the resources and injuries is followed by a description of the plan or program — a set of activities that can be undertaken to hasten the recovery of the natural resources to their baseline conditions. The description of the program includes a discussion of the program's activities, how they will be implemented, and the method of payment. The information about the program coupled with a description of the nature of the resource injuries define the object of choice. The survey then turns to the circumstances of the choice: the disposition of rights to the resources in their baseline conditions, a description of how the choice will be made (*i.e.,* the choice mechanism), and what must be foregone to obtain the object. In the WTP setting of this survey, the respondent faced a simple choice: give up a specified dollar amount in a one-time tax payment in return for the object of choice (a program that will accelerate recovery of the resource) or keep the money and continue to experience the losses associated with the injuries until natural recovery occurs fifty years from now.

The survey uses a referendum as the mechanism to elicit choices. This mechanism provides the respondent with the opportunity to articulate his or her choice by voting *for* or *against* the program, where voting *for* implies getting the program and having the financial burden of paying the tax, while voting *against* implies retention of the money corresponding to the tax amount and continuance of the time profile of injuries associated with natural recovery. This choice mechanism was adopted for a number of reasons.[44] This mechanism is consistent with the

---

[44] The NOAA Panel states (Arrow, p. 4606): "Both experience and logic suggest that responses to open-ended questions will be erratic and biased. However, the referendum format, especially when cast in the willingness to pay mode — `Would you be willing to contribute (or be

mechanisms frequently used in California to decide public issues; and, is familiar and credible for these types of activities. Furthermore, it satisfies the incentive-compatibility conditions required for truthful responses; that is, it meets the condition that respondents evaluate the elements of choice in the same way they would an actual referendum.[45]

Questions that provide information about each respondent's evaluation of the object of choice follow the elicitation of a respondent's choice. They help gauge whether the respondent understood and perceived the information as intended and collect information about the reasons that motivated the reported choices.

Following this series of "debriefing questions," we offered the respondents who voted *for* the program the opportunity to reconsider their decision. Later in the survey, respondents who were still in favor of the program were asked how difficult it would be for them to pay, given the highest tax amount for which they voted, and how strongly they felt about their vote. Respondents indicating that it would be "very difficult", "somewhat difficult", or who said that they were "not sure" how difficult it would be to pay, or indicating that they were "not too

---

taxed) D dollars to cover the cost of avoiding or repairing environmental damage X?' — has many advantages."

[45] A voter has an economic incentive to cast a truthful vote on a ballot proposition as long as he or she thinks the government can provide the object of choice described and that the government will indeed levy the tax stated if the object of choice is provided. The voter is better off voting yes if the object of choice is preferred to paying the tax increase and voting no if it is not.

strongly", or "not at all strongly" in favor of the program, or "not sure" how strongly they favored

the program, were given another opportunity to change their vote. In accordance with

conventional practice, the survey concludes with demographic questions.

§ 4.5 Issues of Reliability Raised by the NOAA Panel

The *Ohio* Court stated,

> On remand, DOI should consider a rule that would permit trustees to derive use
> values for natural resources by summing up all reliably calculated use values,
> however measured, so long as the trustee does not double count. [p. 464]

The Court's statement pointing to the *reliability* of calculated values has stimulated a wide

discussion of the validity[46] one would attach to measures of total value (ILUV) obtained using

CV. As noted in Chapter 1 and above, a panel of experts was formed to help NOAA address the

reliability of CV for damage assessment. In this section, we examine some general issues raised

by the Panel report and discuss briefly other issues raised about the state of CV research and

judgments on its reliability and validity.

§ 4.5.1 Calibration

The Panel's discussion of reliability issues begins with the consideration of the calibration

of CV estimates of total value. The Panel report notes that,

> The contingent valuation method has been criticized for many reasons and the
> Panel believes that a number of these criticisms are particularly compelling.
> Before identifying and discussing these problems, however, it is worth pointing
> out that they all take on added importance in light of the impossibility of
> validating externally the results of CV studies. It should be noticed, however, that
> this same disadvantage must inhere in <u>any</u> method of assessing damages from

---

[46] Bear in mind the distinction between reliability and validity; see footnote 11 in Chapter 1.

deprivation of passive-use. It is not special to the CV approach although, as suggested in section I, there are currently no other methods capable of providing information on these values. [Arrow, p. 4603]

The Panel, seemingly troubled by the fact that it is impossible to externally validate a CV study, referred to the literature on comparisons of hypothetical and real willingness to pay experiments, citing among others, Seip and Strand (1992) and Duffield and Patterson (1991). These two particular studies use a charitable contributions choice mechanism[47], and compare willingness to contribute to particular environmental programs elicited using a "CV like" approach with requests for actual payments. The Panel reports that in the Seip and Strand study "self-reported willingness to pay was significantly greater than `actual' willingness to pay." The Panel then states "These studies suggest that the CV technique is likely to overstate `real' willingness to pay" and "Clearly more experiments would be useful" (Arrow, p. 4604).

---

[47] The results of the studies are consistent with the recognized properties of the charitable contributions format as an inappropriate incentive structure for measuring either "self-reported" willingness to pay or "actual" willingness to pay.

The Panel's call for studies does not explicitly acknowledge the fact that a large set of studies already exists. Indeed, the set of studies that attempt to compare willingness to pay valuations derived from CV surveys with those derived from "actual" market behavior is quite large and the results quite disparate.[48] Depending on the selection of studies providing comparisons, one can conclude that self-reported willingness to pay is less than, greater than, or equal to "actual" willingness to pay. Adding to the potential for confusion is the fact that many, if not all, of the existing studies fail to control for all the elements of choice. For example, it is often the case that the object of choice valued is different across choice situations (*i.e.*, the object of choice may not be the same in the "actual" choice as it was in the "CV" choice). In addition to problems of consistency in the object of choice, many of these studies fail to control for: (1) differences in the institutional setting within which the choice is cast, (2) differences in the informational context used, and (3) differences in the incentive-compatibility attributes of the value elicitation. An important aspect of this last issue concerns whether people accept the implied financial responsibility associated with their CV response(s).

Meaningful comparisons of "actual" and "stated" (or CV) choice results require consistency in the choice elements characterizing each set of decisions. Discrepancies between the two situations call into question comparisons of the WTP estimates derived from each source of choice information. Moreover, there is no reason to expect that ratios of the WTP estimates from "actual" and "stated" choice studies provide meaningful calibration factors for adjusting ILUV estimates obtained from CV studies.

---

[48] See Carson, Flores, Martin, and Wright (1994) for an overview of these studies.

Nothing in economic theory supports the belief that willingness to pay valuations derived from CV surveys produce higher values than actual market transactions when object and circumstances of choice are consistent (*i.e.*, they are for the identical commodity under identical informational, institutional, and incentive-compatible payment schemes). Without theoretical justification for calibration, one may be inclined to turn to the existing body of empirical studies. However, no studies which meet the NOAA Panel guidelines have been conducted that compare CV estimates of ILUV derived from CV studies to comparable studies estimating actual WTP. Without these studies, there is no reason to believe that CV estimates of ILUV would be greater than actual WTP. Moreover, even if a case could be made on the basis of such studies to calibrate a WTP measure of ILUV, there remain questions about such an adjustment. As noted earlier and confirmed by the Panel, WTA-compensation is the appropriate measure of ILUV. WTP already provides a lower bound on WTA.[49] Any arbitrary rule that would lower estimated WTP would lead to a larger discrepancy between WTP and the ideal measure of damages. Thus, it would be improper to "calibrate" downward measures of ILUV obtained from the CV study discussed here.

---

[49] The NOAA Panel states,

> The conceptually correct measure of lost passive-use value for environmental damage that has already occurred is the minimum amount of compensation that each affected individual would be willing to accept. Nevertheless, because of concern that respondents would give unrealistically high answers to such questions, virtually all previous CV studies have described scenarios in which respondents are asked to pay to prevent future occurrences of similar accidents. This is the conservative choice because willingness to accept compensation should exceed willingness to pay, if only trivially; we say more about other biases below [Arrow, p. 4603].

§ 4.5.2 Scope - Inconsistency with Rational Choice

The Panel notes that some CV studies produce results that appear to be inconsistent with

the assumptions of rational choice:

> Usually, though not always, it is reasonable to suppose that more of something regarded as good is better so long as an individual is not satiated. This is in general translated into a willingness to pay somewhat more for more of a good, as judged by the individual. Also, if marginal or incremental willingness to pay for additional amounts does decline with the amount already available, it is usually not reasonable to assume that it declines very abruptly. [Arrow, p. 4603]

> Evidence of this kind has multiplied (see Kahneman and Knetsch, 1992, Desvousges *et al.*, 1992, and Diamond *et al.*, 1992). Desvousges' result is very striking; the average willingness to pay to take measures to prevent 2,000 migratory birds (not endangered species) from dying in oil-filled ponds was as great as that for preventing 20,000 or 200,000 birds from dying. Diminishing marginal willingness to pay for additional protection could be expected to result in some drop. But a drop to zero, especially when the willingness to pay for the first 2,000 birds is certainly not trivial, is hard to explain as the expression of a consistent, rational set of choices. [Arrow, p. 4604]

The Panel's concern over the underlying rationality of values expressed for objects of choice in

CV studies has led to what has become known as a *test for responsiveness to scope*.

It is reasonable to ask: what does responsiveness to scope mean? Unless one is prepared

to make specific assumptions about people's preferences for the object of choice, the only

implication one can draw from rational decision-making is that the economic value of a "large"

amount of the object of choice should not be less than the value of a "small" amount of the same

object. Moreover, what is "large" and "small" is dependent upon individual preferences. As a

result it may not be a simple matter to establish what is "small" and what is "large" even for the

types of goods bought and sold in markets. For example, would an individual considering a

grouping of 40 cans of beer and 2 bottles of wine view that grouping "larger" than a grouping

containing 20 cans of beer and 8 bottles of wine?[50]   Now consider a more intangible object of choice; for example, consider the issue of beach amenities provided by a week's rental of a beach house. Will a larger house located one row back from the ocean provide more beach amenities than a smaller house located in the first row on the ocean? Thus, responsiveness to scope cannot be evaluated without knowing whether people perceive the choices to be *different* and their relative evaluation of large and small.

In its simplest form, a test of scope would vary the nature of the injuries to a natural resource and measure whether the values obtained from a different set of injuries were different from the values obtained for the original set. A comparison of the original set to one described as larger would have to be understood by those surveyed *to be* larger. Similarly, comparisons of the original set of injuries to a smaller set would be expected to yield a reduced WTP for that smaller set of injuries when they are perceived by those surveyed to be smaller. The Panel expresses its desire for such demonstrations by stating,

> . . . some form of internal consistency is the least we would need to feel some confidence that the verbal answers correspond to some reality. [Arrow, p. 4604]

As part of the Lost Use Value Study, we designed and conducted a formal test of scope along the lines noted above. The design of the scope test is discussed in Chapters 5 and 6, and the results are presented in Chapter 10.

---

[50] See Kreps (1990), p. 23.

§ 4.5.3 Sequencing and Substitutes — Implausibility of Responses

The Panel noted a belief maintained by some that individual responses to some CV surveys are implausible (*i.e.*, too large) and that even if the responses seem plausible given the preferences and possible income of the individual, aggregating the values over large populations (as might be the case in estimating total value losses for the purpose of a damage assessment) results in implausibly large values. The Panel states,

> One can envision many possible types of environmental damage -- oil spills or groundwater contamination in many different locations, visibility impairment in a variety of places, and so on. Would the average individual or household <u>really</u> be willing to pay $50 or even $5 to prevent each one? This seems very unlikely, since the total resulting willingness to pay for all such programs could easily become a very large fraction of one's income or perhaps even exceed it.
> In other words, even if the willingness to pay responses to individual environmental insults are correct if only one program is to be considered, they may give overestimates when there are expected to be a large number of environmental problems. [Arrow, p. 4605]

Our response to this concern is straightforward. Chapter 3 provided the rationale for conceptualizing the choice relevant to the measurement of ILUV as a negotiation. Given these choice elements, the appropriate measurement framework for required compensation precludes sequencing. The only legitimate approach for constructing values in a damage assessment is one that focuses on the relevant resources and injuries, and a program to prevent those injuries or to accelerate the recovery of the affected resources to their baseline conditions. Respondents' choices about the program must be informed in a way that would be consistent with negotiation of the terms for required compensation and, as demonstrated in section 3.3.3, would not include sequencing but would include relevant information about substitutes. For example, in this survey, immediately prior to the choice questions, respondents are reminded of the relevant substitutes.

§ 4.5.4 Budget Constraints

The Panel notes that values derived from CV surveys could be influenced by respondents' lack of awareness of their budget constraints, or a perception that their budget is not *meaningful* to their responses. The Panel states,

> Even if respondents in CV surveys take seriously the hypothetical referendum (or other type of) questions being asked them, they may respond without thinking carefully about how much disposable income they have available to allocate to all causes, public and private (see Kemp and Maxwell (1992), for instance). Specifically, respondents might reveal a willingness to pay of, say, $100 for a project that would reduce the risk of an oil spill; but if asked what current or planned expenditures they would forego to pay for the program, they might instead re-evaluate their responses and revise them downward. [Arrow, p. 4605]

There are three points to make with regard to the Panel's concern regarding respondent's cognizance of budget constraints. First, it may be that the concern of the Panel is not with the respondent's awareness of budget constraints *per se*, but a belief that some respondents to CV surveys do not always perceive the financial responsibility implied by their answers. In daily market transactions, economists do not question the awareness of budgets constraints, since it is assumed that consumers are aware of their financial responsibilities. Similarly, if respondents to CV surveys believed they were financially responsible for their answers, specific questions about respondents' awareness of their budget constraints would not be relevant. Accordingly, we placed a high priority on creating a plausible context for respondents to make a decision about whether they would vote to pay the stated tax amount for a program to speed up the recovery of the injured resources in order to ensure that respondents perceived that a personal financial responsibility was associated with their answers. In Chapters 5 and 6, we discuss the features of the survey that serve this purpose.

Second, the Panel provides no empirical evidence to support the notion that consumers

making routine consumption decisions and daily purchases have in mind the category of expenditure that they plan to reduce when they make such purchases. Thus, it seems unreasonable to expect a respondent to a CV survey to be able to describe in detail such explicit consumption trade-offs when deciding how to vote in a CV referendum.

Third, the Panel notes that the issue of budget cognizance is related to,

> . . . the problem identified immediately above where individuals fail to think of the possible multiplicity of environmental projects or policies they might be asked to support. [Arrow, p. 4605]

This view, that at the time respondents face the choice posed by the CV survey, they should have in mind specific requests for payments for public goods other than the services of the injured resource, is inconsistent with WTA-compensation identified in Chapter 3 as the appropriate conceptual framework for natural resource damage assessment.

§ 4.5.5 Information Provision and Acceptance

The Panel notes that CV surveys must provide respondents with information that specifies clearly, and in an understandable way, the nature of the program being offered. The Panel is concerned about the understandability, plausibility, and acceptance of such information by the respondent. The Panel states,

> If CV surveys are to elicit useful information about willingness to pay, respondents must understand exactly what it is they are being asked to value (or vote upon) and must accept the scenario in formulating their responses.

> . . . even when CV surveys provide detailed and accurate information about the effects of the program being valued, respondents must accept that information in making their (hypothetical) choices. If, instead, respondents rely on a set of heuristics ("these environmental accidents are seldom as bad as we're led to believe," or "authorities almost always put too good a face on these things"), in effect they will be answering a different question from that being asked; thus, the resulting values that are elicited will not reliably measure willingness to pay.

(p. 4605)

A great deal of effort was devoted to instrument development to ensure that respondents would understand the information about the speed-up program and its effects. Chapters 5 and 6 describe these efforts.

The Panel also states that respondents "must accept the scenario in formulating their responses" (p. 4605). The Panel's concern is that a lack of acceptance may lead respondents to believe they are choosing an object other than the one the analyst intends to value. While the CV designer should seek to make the information plausible, and therefore believable (Mitchell and Carson, 1989), it is unrealistic and unnecessary to make it an absolute requirement that all respondents accept without question the information provided. For example, it may very well be the case, particularly in a scope test which has to use a version of the injuries that do not accord completely with the facts of the situation, that the information provided is at odds with other information that may be known to the respondent. Sometimes, respondents may have strong beliefs regarding relevant issues (*e.g.*, the ability of ecosystems to recover). In these cases, there is little prospect for presenting alternative evidence contrary to these beliefs that would lead to a complete change in respondents' beliefs.

In order to avoid accepting respondents' choices based on an object other than the one the analyst intends to value, it is necessary to incorporate questions about respondents' level of acceptance of the information presented in the survey so that differences in beliefs can be integrated in the analysis.[51] Questions we used in this survey to collect this information are discussed in Chapter 6.

---

[51] This approach was employed in Carson *et al.*, (1992).

Thus, provided the CV designer has been successful in designing a credible choice, differences in the degree of acceptance of the information describing the object of choice by some respondents should be treated as a natural part of the way people respond to new information. It can be dealt with through the use of a multivariate function describing the factors influencing respondents' choices. Moreover, some lack of acceptance can be helpful in analyzing survey responses. For example, if some respondents believe injuries to a resource were less severe than described in the survey, they should be less willing to vote *for* the restoration program. This is a testable hypothesis that can provide indirect information that can be used in testing the construct validity of the responses. It can produce a conservative bias and, as noted, offers an indirect test of scope.

§ 4.5.6 Extent of the Market

The Panel believes the population of individuals that should be surveyed for damage assessment purposes is determined legally. The Panel states,

> Suits for environmental damages are brought by trustees on behalf of a legally definable group. This group limits the population that is appropriate for determining damages even though individuals outside of this group may suffer loss of passive and active use. [Arrow, p. 4605]

We have no reaction to the Panel's legal interpretation, but simply underscore the Panel's economic judgment that individuals outside of any legally defined group could be willing to make a financial commitment and therefore a choice that would relate to the construction of economic values for the losses due to injuries to natural resources.[52]

---

[52] See Chapter 11 for discussion of this issue in the context of this study.

§ 4.5.7 Warm Glow

The Panel notes a belief held by some that responses to CV choice questions may not reflect the value of the object respondents were asked to consider but rather, the value of "giving to a good cause". The Panel states,

> This has led these critics to conclude that individuals' responses to CV questions serve the same function as charitable contributions -- not only to support the organization in question, but also to feel the "warm glow" that attends donating to worthy causes (see Andreoni, 1989). [Arrow, p. 4605]

To explain observed charitable contributions, Andreoni formulated a specific model to describe people's preferences. He hypothesized that each person's contributions to a charity may well have two separate influences on that individual's level of well-being. The first enhancement would come from some increased level of a public good brought about by the contribution. For example, charitable contributions made to the local volunteer fire department for the purchase of new fire fighting equipment might reduce the anticipated response time and thereby increase the public good known as fire protection which can be enjoyed by all, not just those who made contributions. The second influence does not have a public component and pertains only to the individual making the contribution. For example, I may increase my personal well-being because I simply enjoy giving money to fire fighters and this enjoyment has nothing to do with the enhanced fire protection my contribution may bring about.

Andreoni hypothesized that a dollar of a person's contributions enhances well-being *twice* — once through the public good effect (*e.g.*, increased fire protection) and second through this private effect (*e.g.*, enjoying the act of giving to fire fighters). He described the second effect as *warm glow*. In his model, the decision to contribute was for *one* specific type of charitable

contribution (such as the example of local volunteer fire departments used above), <u>not a contribution to any "good cause"</u> as those who have used his framework out of context have suggested. Since the private component (the second term, as we have described his model above) pertains to a specific contribution, not <u>any</u> contribution, the original Andreoni conception of warm glow has no immediate generalization to all contributions.

Others have broadened Andreoni's original hypothesis regarding things that increase an individual's well-being, to say that <u>any</u> contribution to <u>any</u> charity provides an equivalent private enhancement of well-being (warm glow). This of course could only be true if the contributions to all causes were considered by people to be perfect substitutes for each other. That is, one would be indifferent between making contributions to the local fire department, one's church, or the local SPCA. Casual introspection suggests that the perfect substitution assumption may be suspect. It is our view that the hypotheses regarding preferences and well-being that underlie the generalized model of warm glow (*i.e.*, the model of perfectly substitutable contributions) rely on anecdotes and have not been subject to rigorous empirical tests. Moreover, in this CV study contributions are not elicited; rather, choices are offered at specified tax amounts. The notion that a large number of individuals would receive warm glows from paying taxes or voting to raise their taxes requires one to suspend belief in information about people's expressed concerns over much of the modern history of the state as a taxing authority.

Even if the anecdotes reflect general reality and all the hypotheses regarding the well-being enhancements one derives from making contributions were true (i.e., the warm glow component did exist and all contributions were perfect substitutes), would these hypotheses negate the values the analyst constructs from CV studies? Or, would the hypotheses have any bearing on how one should design the choice situation in a CV study? The answer to both questions is *no*. To

understand why, one need only return to the choice elements identified in the trustee-responsible

party negotiation metaphor developed in Chapter 3. The public is not making a monetary

contribution at all. Rather, the desired valuation concept to be measured is *monetary*

compensation. As noted in Chapter 3, WTP provides a conservative measure of this

compensation. What is relevant to its measurement is the individual's decision to accept financial

responsibility by stating a vote for the proposed program. There is no contribution involved and

thus warm glow is irrelevant.

§ 4.6 NOAA Panel Survey Design and Administration Guidelines

The report of the NOAA Panel describes a detailed set of survey design and administration

guidelines. Some of these are novel suggestions. The majority are standard practice for

conducting any high quality survey research and can be found in high quality CV surveys. The

Panel notes,

> In this section we try to lay down a fairly complete set of guidelines compliance
> with which would define an ideal CV survey. A CV survey does not have to meet
> each of these guidelines fully in order to qualify as a source of reliable
> information to a damage assessment process. Many departures from the
> guidelines or even a single serious deviation would, however, suggest
> unreliability *prima facie*. [Arrow, p. 4608]

We have carefully examined each of the Panel's suggestions, many of which are consistent with

our standard practice (see Carson *et al.*, 1992). When the suggestion represents a novel

departure from this standard practice, an examination of the proposal was conducted as part of

our instrument design. These particular instances concerned recommendations where we could

find little theoretical or empirical support in the literature. On the basis of that examination,

which involved a major development effort to test the assumptions which underlie several of the

recommendations, we have chosen in some instances to differ with the specific procedure recommended by the Panel. In every case, however, we paid careful attention in designing the survey for this study to address what we interpret as the Panel's underlying concern.

§ 4.6.1 Summary of Panel Guidelines

In this section we list the 23 survey design guidelines that are potentially applicable to a CV survey such as this one.[53] We implemented 20 of these guidelines, a few of them with qualifications. Of the others, we believe there are good reasons why two of them do not need to be implemented and the third is not applicable to our survey. Listed below are the specific NOAA Panel recommendations implemented in this CV survey and the text (taken verbatim from the NOAA Panel report) describing the nature of each. In the following section we present the reasoning behind each of the recommendations we do not implement or that we implement with qualifications.

> **Conservative Design — Implemented:** Generally, when aspects of the survey design and the analysis of the responses are ambiguous, the option that tends to underestimate willingness to pay is preferred. A conservative design increases the reliability of the estimate by eliminating extreme responses that can enlarge estimated values wildly and implausibly.

Elicitation Format — Implemented:

> The willingness to pay format should be used instead of the compensation required because the former is the conservative choice.

---

[53] Two of Panel guidelines do not apply to survey design issues. These are a recommendation that critical features of a CV survey be pre-approved by both sides in a legal action and the government undertake the task of creating a set of reliable reference surveys.

Referendum Format — Implemented:

The valuation question should be posed as a vote on a referendum.

Personal Interview — Implemented:

The Panel believes it unlikely that reliable estimates of values could be elicited with mail surveys. Face-to-face interviews are usually preferable, although telephone interviews have some advantages in terms of cost and centralized supervision.

Sample Type and Size — Implemented:

Probability sampling is essential for a survey used for damage assessment.[54] The choice of sample specific design and size is a difficult, technical question that requires the guidance of a professional sampling statistician.

Minimize Nonresponses — Implemented:

High nonresponse rates would make the survey results unreliable.

---

[54] The following footnote is taken directly from the Panel report: "This need not preclude use of less adequate samples, including quota or even convenience samples, for preliminary testing of specific experimental variations, so long as order of magnitude differences rather than univariate results are the focus. Even then, obvious sources of bias should be avoided (*e.g.*, college students are probably too different in age and education from the heterogeneous adult population to provide a trustworthy basis for wider generalization)".

Reporting — Implemented:

Every report of a CV study should make clear the definition of the population sampled, the sampling frame used, the sample size, the overall sample non-response rate and its components (e.g., refusals), and item non-response on all important questions. The report should also reproduce the exact wording and sequence of the questionnaire and of other communications to respondents (e.g., advance letters). All data from the study should be archived and made available to interested parties (see Carson et al., (1992), for an example of good practice in inclusion of questionnaire and related details; as of this date, however, the report has not been available publicly and the data have not been archived for open use by other scholars).

Cross-tabulations — Implemented:

The survey should include a variety of other questions that help to interpret the responses to the primary valuation question. The final report should include summaries of willingness to pay broken down by these categories. Among the items that would be helpful in interpreting the responses are:

> Income,
> Prior Knowledge of the Site,
> Prior Interest in the Site (Visitation Rates),
> Attitudes Toward the Environment,
> Attitudes Toward Big Business,
> Distance to the Site,
> Understanding of the Task,
> Belief in the Scenarios, and
> Ability/Willingness to Perform the Task.

Accurate Description of the Program or Policy — Implemented:

Adequate information must be provided to respondents about the environmental program that is offered. It must be defined in a way that is relevant to damage assessment.

Adequate Time Lapse from the Accident — Implemented:

The survey must be conducted at a time sufficiently distant from the date of the environmental insult that respondents regard the scenario of complete restoration as plausible. Questions should be included to determine the state of subjects' beliefs regarding restoration probabilities.

Careful Pretesting of a CV Questionnaire — Implemented:

Respondents in a CV survey are ordinarily presented with a good deal of new and often technical information, well beyond what is typical in most surveys. This requires very careful pilot work and pretesting, plus evidence from the final survey that respondents understood and accepted the main description and questioning reasonably well.

Checks on Understanding and Acceptance — Implemented:

The above guidelines must be satisfied without making the instrument so complex that it poses tasks that are beyond the ability or interest level of many participants.

Yes/No Follow-ups — Implemented:

Yes and no responses should be followed up by the open-ended question: "Why did you vote yes/no?" Answers should be carefully coded to show the types of responses, for example: (i) It is (or isn't) worth it; (ii) Don't know; or (iii) The oil companies should pay.

Pretesting of Photographs — Implemented:

The effects of photographs on subjects must be carefully explored.

Pretesting for Interviewer Effects — Implemented:

An important respect in which CV surveys differ from actual referenda is the presence of an interviewer (except in the case of mail surveys). It is possible that interviewers contribute to "social desirability" bias, since preserving the environment is widely viewed as something positive. In order to test this possibility, major CV studies should incorporate experiments that assess interviewer effects.

Reminder of Undamaged Substitute Commodities — Implemented:

Respondents must be reminded of substitute commodities, such as other comparable natural resources or the future state of the same natural resource. This reminder should be introduced forcefully and directly prior to the main valuation question to assure that respondents have the alternatives clearly in mind.

Present Value Calculations of Interim Losses — Implemented with Qualification:

It should be demonstrated that, in revealing values, respondents are adequately sensitive to the timing of the restoration process.

Deflection of Transaction Value — Implemented with Qualification:

The survey should be designed to deflect the general "warm-glow" of giving or the dislike of "big business" away from the specific environmental program that is being evaluated. It is possible that the referendum format limits the "warm glow" effect, but until this is clear the survey design should explicitly address this problem.

Alternative Expenditure Possibilities — Implemented with Qualification:

Respondents must be reminded that their willingness to pay for the environmental program in question would reduce their expenditures for private goods or other public goods. This reminder should be more than perfunctory, but less than overwhelming. The goal is to induce respondents to keep in mind other likely expenditures, including those on other environmental goods, when evaluating the main scenario.

Burden of Proof — Implemented with Qualification:

Until such time as there is a set of reliable reference surveys, the burden of proof of reliability must rest on the survey designers. They must show through pretesting or other experiments that their survey does not suffer from the problems that these guidelines are intended to avoid. Specifically, 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 instrument or to the valuation question.
• 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 referendum that are not followed up or explained by making reference to the cost and/or the value of the program.

Temporal Averaging — Not Implemented:

Time dependent measurement noise should be reduced by averaging across independently drawn samples taken at different points in time. A clear and substantial time trend in the responses would cast doubt on the "reliability" of the finding.

"No-answer" Option — Not Implemented:

A "no-answer" option should be explicitly allowed in addition to the "yes" and "no" vote options on the main valuation (referendum) question. Respondents who choose the "no-answer" option should be asked nondirectively to explain their choice. Answers should be carefully coded to show the types of responses, for example: (i) rough indifference between a yes and a no vote; (ii) inability to make a decision without more time or more information; (iii) preference for some other mechanism for making this decision; and (iv) bored by this survey and anxious to end it as quickly as possible.

Steady State or Interim Losses — Not Applicable:

It should be made apparent that respondents can distinguish interim from steady-state losses.

§ 4.6.2 Discussion of Panel Guidelines

In this section we discuss the seven recommendations which we did not implement or implemented with qualification.

§ 4.6.2.1 Steady State or Interim Losses — Not Applicable

The recommendation that "It should be made apparent that respondents can distinguish interim from steady-state losses" is not applicable to the Lost Use Value Study. It is not entirely clear how the Panel implicitly defined *interim* and *steady state* losses, but the following quote from the Panel's report provides some guidance:

Typically, environmental damages from oil spills or similar accidents are severe

for some period of time -- weeks, months, or sometimes a few years -- and gradually are reduced by natural forces and human efforts to a low or possibly even zero steady state level. In some circumstances, passive-use losses derive only or mostly from the steady state conditions; thus, if passive use value derives from species diversity, even a considerable loss of birds or mammals which does not endanger the species will give rise to no loss in value. [Arrow, p. 4608]

We conclude from the above statement that the Panel associates *interim* losses with injuries of short duration, for example, injuries that last for only a few years, while *steady state* losses are associated with injuries that continue for periods of time in excess of a few years.

We use italics for the Panel's expression *interim* losses to underscore the fact that the definition of *interim* loss we are attributing to the Panel is not coincident with the definition of interim lost use normally employed in damage assessment and employed in the Lost Use Value Study. In damage assessment, interim lost use is not time-qualified; that is, it does not pertain exclusively to injuries of a short duration, but is the term used to define losses due to injuries during the interim from onset of injury to complete recovery of the resources to what would be their baseline conditions.

As noted in Chapter 2, the releases in question and associated injuries began more than 40 years ago; and conservative estimates of natural recovery suggest that a return to baseline conditions is well into the next century. Therefore, there is no reason to believe that these are short duration losses that the Panel would classify as *interim*. Since there are no *interim* losses as defined by the Panel, there is no reason to query respondents about such losses.

§ 4.6.2.2 Present Value Calculations of Interim Losses — Qualification

The Panel suggests that respondents should be adequately sensitive to the timing of the restoration process. The context in which this recommendation is placed implies that the Panel

was considering the *short-term* effects of oil spills and the potential for a fairly rapid recovery.

For example, in introducing this issue, the Panel observes:

> Typically, environmental damages from oil spills or similar accidents are severe for some period of time — weeks, months, or sometimes a few years — and gradually are reduced by natural forces and human efforts to a low or possibly even zero steady state level. ... CV surveys accordingly have to be carefully designed to allow respondents to differentiate interim from steady state passive-use loss, and, if there is interim passive-use loss, to report its present value correctly. [Arrow, p. 4608]

This explanation makes clear that to the extent the Panel's concerns are relevant to our CV questions, they are satisfied by the framework. There are two reasons for this conclusion. First, our objective is to measure prospective ILUV, assuming full restoration. When a respondent chooses the proposed program, there are no steady state passive use losses. Second, the injuries associated with the Southern California Bight do not conform to the "short-term" oil spill framework envisioned by the Panel's guidance.

§ 4.6.2.3 Deflection of Transaction Value — Qualification

As noted earlier in this chapter (section 4.5.7), in the absence of any rigorous tests we believe the concept of "warm glow" as developed by Andreoni is irrelevant to a CV survey such as the one described here. What is relevant in this context is the Panel's suggestion that the respondent's attention should be focused on the specific injuries to natural resources and away from expressions of general preferences for improving the environment. We placed a high priority on this requirement in designing the instrument for the Lost Use Value Study. Two features in particular — the use of the referendum format and an income tax payment vehicle — serve to

---

Furthermore, the panel's framework may not apply to all or any oil spills either. In any event, whether the short-term "oil spill" framework of the Panel applies is an empirical issue.

enhance the realism of the choice situation and therefore help to deflect any transaction value.

§ 4.6.2.4 Burden of Proof: Limitations on Yes/No Follow-up Questions — Qualification

The Panel calls for open-ended inquiries to allow respondents to explain their reasoning in answering the CV choice questions. Their recommendation must be considered in light of the literature in psychology on the reliability of introspection questions that ask respondents to explain how they arrived at their reported attitudes. This research indicates that although people generally have good insights into their likes and dislikes and can report those attitudes well, the process underlying their thinking is more difficult to elicit. The literature suggests three guidelines for efforts to collect this information. All three were incorporated in our main study questionnaire when respondents were asked about the reasons for their choices.

First, requests for explanations of choices should be treated as containing "traces" (not specifics) of the cognitive processes leading to a respondent's decision. In short, one should not expect to obtain a detailed, fully accurate explanation of all the reasons why an individual made a particular choice. Second, these introspective questions should be placed after all important choice questions, because some literature indicated they can be disruptive. Finally, no attempt should be made to request a respondent's view of his or her choice in comparison with the choices that would be made by others.

§ 4.6.2.5 Alternative Expenditure Possibilities — Qualification

In its recommendations, the Panel says,

Respondents must be reminded that their willingness to pay for the environmental program in question would reduce their expenditures for private goods or other public goods. This reminder should be more than perfunctory, but less than overwhelming. The goal is to induce respondents to keep in mind

other likely expenditures, including those on other environmental goods, when evaluating the main scenario. [Arrow, p. 4609]

Insuring that respondents understand and consider in their deliberations the consequences of the choices they make in response to a CV question is our standard CV practice. We believe this quote reflects the Panel's concern with this same understanding and consideration of consequences. In the current study, we emphasized to the respondent the financial responsibility associated with votes *for* the program at a point immediately before the voting choice questions and gave the respondent an opportunity to reconsider his or her vote at a later point in the survey after further emphasizing the respondent's financial responsibility implied by the choice.

§ 4.6.2.6 Temporal Averaging — Not Implemented

The Panel suggested that "time dependent measurement noise should be reduced by averaging across independently drawn samples taken at different points in time." One might interpret the Panel's proposal to mean that the identical final CV survey (*i.e.*, the survey as it stands after all design work is completed and pilot testing accomplished) should be administered to a random sample of the target population on at least two occasions, separated from each other by an unspecified period of time. Estimates of lost total value calculated from these surveys would then be averaged. On the basis of specific instrument development work described immediately below, which found no evidence for "time dependent measurement noise" for CV surveys with design characteristics similar to the present one, we did not implement this recommendation.

We replicated the Exxon Valdez oil spill (EVOS) CV survey more than two years after it

---

See Chapter 6 for further details.

was originally fielded. The Exxon Valdez study, reported in Carson *et al.,* (1992), was reviewed by the Panel and used by them as an example in describing several of the key elements in their recommendations. More generally, the spill represented a large disaster at a particular point in time, an occurrence that is likely to be characteristic of the types of incidents that the Panel felt would benefit from temporal averaging. Because the EVOS instrument is closely comparable to the present questionnaire in its design and implementation, it was an appropriate vehicle to use to examine the issue of time dependent measurement noise and several other Panel suggestions.

The EVOS replication was conducted for us as part of the instrument development for the Lost Use Value Study by the National Opinion Research Center (NORC) of the University of Chicago in May-July, 1993. The interviews were conducted in-person with a probability sample of adults chosen from 34 counties throughout the United States. An empirical test of the hypothesis that the pattern of votes *for* and *against* a program to prevent a future "Exxon Valdez" type oil spill in Prince William Sound has changed over the two year period between the first and the second administrations of the EVOS instrument is presented below. On the basis of these results we find no empirical support for the recommendation to temporally average results from our current study.

The Exxon Valdez CV instrument has the same structure as the present survey: general attitude questions at the beginning, description of injuries and then a program that would prevent them, referendum format, tax payment vehicle, for/against responses to different tax amounts, follow-up questions, and opportunities for the respondents to change their votes. Like the present

---

The questionnaire was slightly amended to reflect the change in the timing of the survey in relation to the oil spill.

NORC is a nationally recognized professional survey research organization. Chapter 5 and Appendix B.1 contain further discussion of the NORC survey and its findings.

survey, it was conducted in-person by professional interviewers. The specific text read to the

respondents was as follows:

> Because everyone would bear <u>part</u> of the cost, we are using this survey to ask people how they would vote if they had the chance to vote on the program.
>
> We have found some people would vote <u>for</u> the program and others would vote <u>against</u> it. Both have good reasons for why they would vote that way.
>
> Those who vote <u>for</u> say it is worth money to them to prevent the damage from another large spill in Prince William Sound.
>
> Those who vote <u>against</u> mention concerns like the following:
>
> Some mention that it won't protect any other part of the country except the area around Prince William Sound.
>
> Some say that if they pay for this program they would have less money to use for other things that are more important to them.
>
> And some say the money they would have to pay for the program is more than they can afford. (PAUSE)
>
> Of course whether people would vote for or against the escort ship program depends on how much it will cost <u>their household</u>.
>
> <u>At present</u>, government officials estimate the program will cost <u>your</u> household a total of $___. You would pay this in a special one time charge in addition to your regular federal taxes. This money would <u>only</u> be used for the program to prevent damage from another large oil spill in Prince William Sound. **(PAUSE)**
>
> <u>If</u> the program cost your household a total of $___, would you vote . . . **(READ CATEGORIES AND CODE ONE ONLY)**
>
>> For the program, ...1
>> Against the program,   2
>>
>> NOT SURE 8

The blank tax amounts ($__) were randomly assigned to one of four dollar amounts ($10, $30,

$60, or $120).

The most direct test of the influence of time (*i.e.*, the lapse between the original EVOS

survey in the late winter of 1991 and the replication using the NORC survey in the summer of

1993) is provided by evaluating whether time influenced the distribution of *for* and *against*

choices in responding to the offered program. In 1991, 51.5 percent of the sample (N=1,043)

voted for the oil spill prevention program compared with 52.7 percent of the 1993 NORC sample

(N=300). This difference is not statistically different (p=0.714). Table 4.1 reports the distribution

of *for* and *against* votes at each of the four tax amounts used in the EVOS instrument. For each

amount, the differences between the EVOS responses and the NORC responses were not

statistically significant.

**Table 4.1 Comparison of Votes at Different Tax Amounts
for the EVOS and NORC CV Surveys**

CHOICE	EVOS [*1/91—4/91*]	NORC [*5/93—7/93*]
Tax = $10	(N=264)	(N=87)
Vote For	67.4%	67.8%
Vote Against	32.6%	32.2%
	χ^2=0.005; p=0.946	
Tax = $30	(N=267)	(N=66)
Vote For	51.7%	56.1%
Vote Against	48.3%	43.9%
	χ^2=0.406; p=0.524	
Tax = $60	(N=255)	(N=81)
Vote For	50.6%	49.4%

The p-value is a simple value used to describe the test results. As a rule, hypotheses tests adopt a specific significance level (often 5 percent). This significance level specifies the probability of incorrectly rejecting a "true" null hypothesis. Thus with the selection of 5 percent significance level, one would be accepting the chance of making mistakes 5 percent of the time this test was repeated with exactly the same hypotheses and type of information. The p-value computes what the level of significance would have to be adopted to reject the hypothesis.

In all cases, the null hypothesis of comparable distribution of responses cannot be rejected at the conventional levels of confidence.

Vote Against	49.4%	50.6%
χ^2=0.036; p=0.850		
Tax = $120	(N=257)	(N=66)
Vote For	34.2%	33.3%
Vote Against	65.8%	66.7%
χ^2=0.019; p=0.890		

§ 4.6.2.7 No-Answer Option — Not Implemented

The NOAA Panel included a recommendation that CV surveys explicitly offer respondents a third, "no-answer" option. In its rationale for this recommendation, the Panel points out that in national split-sample experiments, large numbers of people take the "don't know" option when it is offered as an answer option to typical attitude questions. The Panel was concerned that there may be a comparable percent of respondents in contingent valuation surveys who give WTP responses when forced to do so but whose answers do not reflect meaningful opinions on the issue. In addition, the Panel suggested that an explicit would-not-vote option in a contingent valuation instrument would better simulate real referenda where voters always have the opportunity of not voting.

Implementing the "would-not-vote" recommendation in the contingent valuation context has a potentially serious cost: the loss of choice information from a portion of the sample. An alternate view to the one expressed by the Panel holds that this sacrifice is not necessary, because most or all of those who take an offered "would-not-vote" option are in fact capable of making a meaningful voting decision. This view holds that offering the would-not-vote option encourages respondents to "satisfice" rather than to expend the effort necessary to give considered responses.

In a CV interview, by the time people reach the point of voting, they have received a great

---

See Appendix B.1 for a discussion of this perspective.

deal of information about the issue and most are likely to be able to make a decision one way or the other if the study is well designed. In such a study, if a person cannot make a decision, he or she is not pressured to do so by a CV interviewer, who is instructed to accept "not sure" answers whenever they are offered to a voting question.

We used a split-sample design with the NORC survey described earlier to examine whether the lack of a would-not-vote option biases the findings of a CV survey in the ways the Panel suggested. This type of test, where one random sub-sample receives one treatment and another random sub-sample receives a different treatment, is a standard procedure used by survey researchers to determine whether variations in question wording or context affect responses (Schuman and Presser, 1981; Turner and Martin, 1984). Professional interviewers from NORC administered four versions of the EVOS instrument. Respondents were assigned randomly to the four treatments. Here we compare two of these treatments: the standard version of the EVOS instrument and a would-not-vote version. The standard version offered only the for/against options; the would-not-vote version was identical to the standard version in every way except at the willingness-to-pay questions where it explicitly offered a would-not-vote option in addition to the for/against options. In both treatments, interviewers accepted "not sure" responses if the respondent expressed this point of view.

This test yielded three main findings. First, the percent who took the would-not-vote

---

Sometimes called a "split-ballot" test.

[55] We used the EVOS instrument because it used a design very similar to the present survey and was fully field-tested and ready to administer whereas the instrument for this study was still in the development stage. The only modifications we made in the Alaska instrument were those required by the tests.

This is the version we used earlier to compare with the original EVOS findings to test the need for temporal averaging.

option when offered was 9.3 percent. This is considerably lower than the average of about 25 percent expected by the NOAA Panel. An additional 8.4 percent in the would-not-vote version said they were "not sure" how they would vote, which is close to the 6.7 percent who said they were "not sure" in the standard version.

Second, when those respondents who chose the would-not-vote option were counted as voting against the program (a conservative assignment), the two treatments resulted in virtually identical voting patterns. One comparison, displayed in Table 4.2, shows that the proportion of the respondents voting in favor of the program to protect Prince William Sound was nearly identical regardless of whether or not the would-not-vote option was offered. It appears that virtually all of those who take the would-not-vote option would otherwise have voted against the program.

Third, to explore whether offering the would-not-vote option improved data quality by eliminating respondents who lacked meaningful opinions, we assessed how well we could predict respondents' choices (*i.e.,* a vote *for* or a vote *against*) using their attitudes and beliefs (*e.g.,* how effectively they felt the escort ship program prevented oil spills, how much they supported programs to protect wilderness areas), demographic characteristics such as income, and characteristics of the choice they were given (*e.g.*, the amount of tax their household would have to pay). If omitting the would-not-vote option led some respondents to select vote *for* or vote *against* choices in a haphazard manner, then voting decisions should be predicted less well by these various factors than when the would-not-vote option is included. This turned out not to be the case: the set of predictors explained the choices just as well when the would-not-vote option was offered as when it was omitted. A complete discussion of this analysis is contained in Appendix B.1.

Table 4.2 Effects of Not Vote Option — Composite Across Tax Amounts

Choice	Standard Version	Would Not Vote Option Offered
Tax = $10	(N=87)	(N=82)
Vote For	67.8%	73.2%
Vote Against	32.2%	26.8%
χ^2=0.581; p=0.446		
Tax = $30	(N=66)	(N=87)
Vote For	56.1%	49.4%
Vote Against	43.9%	50.6%
χ^2=0.662; p=0.416		
Tax = $60	(N=81)	(N=73)
Vote For	49.4%	45.2%
Vote Against	50.6%	54.8%
χ^2=0.269; p=0.604		
Tax = $120	(N=66)	(N=80)
Vote For	33.3%	38.8%
Vote Against	66.7%	61.3%
χ^2=0.459; p=0.498		

This result is inconsistent with the Panel's logic and consistent with an alternative position that offering the would-not-vote option is undesirable because it encourages respondents who would otherwise vote against the program to take an easy out and accept the would-not-vote option. If our test had found that offering the would-not-vote option significantly lowered CV willingness-to-pay values or improved the quality of the data we would have used that format in this study. Because it did not, we concluded that there were no grounds for following the Panel's recommendation in this particular case.

§ 5 Development of the Survey Instruments

§ 5.1 Introduction

The survey instrument for the main study was developed over 31 months, beginning in August 1991 and ending in March 1994, when the final instrument was put into the field. During this development period, the NOAA Panel released its report. While many of the considerations posed by that report reflected our standard practice, by endorsing certain design options, the report mandated serious consideration of several other issues. Chief among these was sensitivity to scope for which we developed a second survey instrument, identical to the first except for the injury description and other wording dependent on the injury description. We refer to the first instrument as the *base* version and to the second as the *scope* version. The final versions of these two instruments were used to examine whether respondents' choices were sensitive to the size of the injury. The discussion that follows focuses on the development of the base instrument, unless otherwise specified.

As discussed in Chapter 4, the central portion of the survey instrument describes the elements of the choice including the description of the object of choice and circumstances of the choice. The mechanism by which each respondent was given the opportunity to state a choice was a referendum where the respondent was asked to vote *for* or *against* a program (*i.e.,* the object of choice) to speed up the affected species' recovery from the described effects of DDT and PCB's. Given the elements of choice described in the survey, respondents' decisions can be used

---

See Carson *et al.*, (1992).

The scope version (see Appendix A-2) also includes an extra question, C-23, that asks respondents if they would consider the problem caused by the two chemicals to be more serious if Bald Eagles and Peregrine Falcons were also affected.

See Chapter 3 for a discussion of the choice framework.

to construct a measure of prospective interim lost use value (ILUV). Other questions preceding and following the presentation of the choice questions ask about respondent attitudes, familiarity with the chemicals and the deposit, understanding of the choice elements, and personal characteristics. During the interview, show cards, maps, and diagrams are shown to respondents to help convey the information presented verbally by the interviewers.

In this chapter, we discuss the development of the main study survey instruments with a particular focus on the development of a credible description of the object of choice and circumstances of the choice. Throughout this development process, we followed the basic objectives outlined in section 4.3 of Chapter 4 and those discussed below.

§ 5.2   Objectives of the Instrument Development

We conducted an extensive program of instrument development for this study. The first stage involved exploratory work, primarily through focus groups. In the next stage, a first draft of the questionnaire was continually revised while testing it in a series of cognitive interviews followed by several small field pretests. During the third stage we conducted a series of pilot surveys, as well as additional cognitive interviews and several more pretests. The draft instrument was peer reviewed during this third stage by specialists in information design theory, resource economics, psychology, and survey research. Throughout each stage of this process, we followed established survey research practices to ensure the reliability of the final results. Later in our

---

Reproductions of the graphic sets can be found in Appendices A-1 (base) and A-2 (scope).

See Appendices A-1 and A-2 for copies of the base and scope survey instruments, respectively.

[56] The reader is reminded that in this report we use the term "reliability" in the legal sense of "dependable", "trustworthy". This is similar to the term "validity" as it is used in the survey research literature (*e.g.,* Mitchell and Carson, 1989).

work, we conducted a similar development program for an instrument for the scope test.

In the development process we sought to have the instrument meet the following objectives: the instrument should be

1.   consistent with economic theory;
2.   comprehensible to respondents;
3.   focused on the set of defined injuries;
4.   plausible in regard to the choice mechanism; and
5.   perceived overall as neutral by the respondents.

The first objective was to develop an instrument that was consistent with the economic theory outlined in Chapter 3. Specifically, the instrument was designed to enable a monetary measure of economic value to be constructed from a well-defined choice regarding the specified set of natural resource injuries. Further, as Chapter 3 discussed, even though the ideal elements of choice cannot always be implemented in practice, the formulation of these elements should nonetheless mimic as closely as possible the condition appropriate to the compensation required by the public to permit the temporal pattern of injuries.

The second objective is a basic survey research goal — that respondents from all educational levels and varied life experiences comprehend the language, concepts, and questions used in the survey.

The third objective was to have the respondents focus on only the defined set of injuries. This objective required carefully describing the specific injuries and their recovery time in such a way as to minimize the possibility that respondents would envision a more extensive or less extensive set of injuries. The presentation of the injuries was guided by our findings during instrument development. We used open-ended debriefing questions and follow-up questions to

_____

  As noted in Chapter 2, the set of injuries to be valued was provided by Trustees representatives and the injuries as described in the final survey instruments were reviewed by the Trustees prior to the fielding of the main study.

monitor our success in meeting this goal. Follow-up questions were used to evaluate the relationship between respondent choices and those instances when the respondent apparently envisioned injuries which differed in some way from the set described in the instrument. As noted above, we also tested whether respondents were sensitive to the size of the injury by administering two versions of the instrument to split samples.

Our fourth objective was to design a plausible choice mechanism. Even if a respondent understands the choice, he or she will not take it seriously if it is not plausible. To this end, we used a referendum mechanism: each respondent was asked to make a decision as to whether he or she would vote for or against a program that, if adopted, would cost his/her household a specified amount in addition to what the household already pays for other public goods and household expenses. A large number of other design decisions to enhance plausibility will be noted in this and the following chapter. For example, describing the State as the sponsor helped enhance the referendum's realism and the State's intent in conducting the survey was explained in such a way that respondents would find it reasonable to be asked about how they would vote given the particular set of injuries described to them.

Perceived neutrality was the fifth goal: respondents should not perceive the purpose of the interview as the State's promotion of a particular choice. To this end, we took care to avoid bias in the wording and the sequence of the material, and we encouraged respondents to consider a number of reasons why they might not want to vote for the program.

In addition to the objectives discussed above, we followed a conservative strategy when faced with instrument design choices where there was no apparent correct choice based on the facts, theory, methodological considerations, or the recommendations of the NOAA Panel report.

See Section 5.7 for a description of the *scope* version.

In these cases, we chose the design alternative that, if it had any effect on the respondent, would tend to reduce the likelihood of a vote *for* the program.

§ 5.3 Instrument Design — Stage 1

The design work for the survey began in August of 1991 with a series of five focus groups in different locations throughout California. The location and dates of these five groups are listed in Table 5.1. The focus group sites were used to recruit participants from several different areas in Los Angeles, San Diego, and San Jose.

Table 5.1 Stage 1 Focus Groups

Focus Group	Location in CA	Date Conducted
1	Burbank/San Fernando Valley	August 6, 1991
2	Torrance	August 7, 1991
3	San Diego	August 19, 1991
4	San Jose	August 20, 1991
5	Torrance	August 21, 1991

Focus groups are group discussions, lasting, in our case, about two hours in length. The group discusses topics introduced by a moderator who leads the discussion (Greenbaum, 1993). The goal of the discussion is to obtain information from the participants. Focus groups have often been used to learn about how people think about commercial products or political candidates.

Focus groups were just one of several techniques used during instrument development.

We conducted an additional set of nine focus groups in 1993 (see Table 5.7) during the design of the scope instrument.

The two Torrance groups were recruited from different areas, one south of the facility and the other north.

They are also used to improve survey design. Although those who choose to participate in focus groups are not a random sample of the public, information learned from the groups can be checked later in the instrument development process by conducting pretest and pilot interviews in the field.

In the case of designing contingent valuation questionnaires, focus groups offer the opportunity in the early stages of design to explore participants' beliefs, attitudes, and knowledge about the survey's subject matter, and to obtain the participants' reactions to possible scenario elements (Desvousges, Smith, Brown, and Pate, 1984). For this study we wanted to learn what knowledge respondents might have about the particular chemical deposit and its effect; what beliefs they held that might affect their responses; and how plausible they found possible elements of the choice we could use in the questionnaire.

The focus groups for this study were conducted in facilities designed for focus group research. For example, all of the facilities had an observation room where researchers could discreetly observe the discussion through a one-way mirror. Eight to 12 participants were randomly recruited by the focus group facility staff either from their own databases or randomly from local telephone directories. For their time, the participants were paid the standard fee recommended by the facility .

We provided the recruiters with a screening questionnaire to recruit people in certain age, education, and sex categories and to filter out any persons who had previously taken part in any focus group. We typically used quotas to ensure that the group included a balanced number of men and women, a range of ages, and a range of educational attainments. To reduce the chance that those who agreed to participate were especially interested in the discussion topic, the screener described the purpose of the group in general terms: "On (date), we are holding a group

discussion to gather area resident opinions on a current state public issue."

So that we could learn the participants' pre-existing views about the subject matter of the survey, the specific subject matter—the DDT/PCB deposit and its effects—was only revealed, later in the session, after an initial discussion. The conversations were tape-recorded, and the audio tapes were transcribed for further analysis.

Most focus group participants had heard of DDT and some of PCB's; most had not heard of the particular DDT/PCB deposit being discussed. The wildlife injuries were plausible to many, as was the concept of bio-magnification up the food chain. There was a rather widespread knowledge that DDT causes eggshell thinning. The idea of a program to cover the contaminated sediment showed promise of being plausible, but various concerns were raised that would have to be addressed in the survey instrument, such as a fear that it would stir up the sediment, a concern that it would not be effective, distrust of the State's ability to carry it out, and a desire to know whether something like this had been done elsewhere. Some participants made comments indicating that they would not be willing to pay for this for a variety of reasons, including a belief that taxes were too high already and that the affected species would eventually recover naturally. Hearing about the effects of DDT and PCB's in fish raised concerns in some participants' minds that the chemicals posed a health threat to humans.

§ 5.4 Key Design Issues

The focus groups helped us to assess the instrument design we had in mind, what information was important to present during the interview, and which potential sources of misunderstanding required addressing in the instrument's wording to avoid biasing the findings. On the basis of this information and decisions about a number of important design issues, the

initial draft of the survey instrument was developed. These design issues involved the specific choice mechanism, the elicitation method, the nature of the payment vehicle, the number of years over which payments would be collected, the sequence in which the choice elements were presented, whether to offer respondents a specific "not vote" alternative in asking the willingness-to-pay question, whether to present the damages as part of a sequence with other goods, the choice of substitutes, quantity of information presented, and visual aids. These decisions were subjected to peer review at several points during the design process. A brief discussion of these decisions follows.

Choice mechanism. We framed the choice for the respondent as a referendum voting decision where the respondent was asked to state how he or she would vote on a well-defined object of choice at a specified tax amount. In a national contingent valuation survey (Carson *et al.*, 1992) and in the present study, we have found most respondents easily comprehend a referendum vote decision and take this type of voting question seriously. Voting on ballot propositions concerning government policies has a long history in American politics. In California propositions are frequently placed on the ballot at the state and local level in California. Political mechanisms of this kind have desirable theoretical attributes that CV surveys are well suited to realize since they are able to provide key information about the good and its provision in a controlled setting that optimizes respondent attention and comprehension. The referendum model has been widely adopted by CV practitioners and, as noted in Chapter 4, was endorsed by the NOAA Panel (Arrow *et al.*, 1993; p. 4608).

---

[57] Some design decisions were made at later stages in the instrument development process. The decisions discussed here all pertain to the final instrument.

[58] See Mitchell and Carson (1989; pp. 94-97) for a discussion of the relationship between a CV referendum and actual referenda on this point.

Elicitation method. We used a binary, discrete-choice elicitation question which states a tax amount and then offers respondents the choice to vote for or against the program. This type of question is easier for respondents to answer than an open-ended question that asks them for the exact amount they are willing to pay for the program (*i.e.,* object of choice). Furthermore, people usually make decisions for most goods they purchase in this take-it-or-leave-it manner. The referendum context generally provides respondents with an incentive to vote *for* if they would rather implement the program and pay the amount specified and to vote *against* if they would rather not pay the amount specified.

A second, binary discrete-choice question followed the first. Those who initially voted *against* were asked to make a decision about a lower amount, and those who initially voted *for* were asked to make a decision about a higher amount.

Payment vehicle. The payment vehicle specifies how the respondent would pay for the object of choice. The link between the payment vehicle and the object must be plausible and credible, and it should bring the relevant budget constraint to mind. We chose our payment vehicle—an additional amount on the respondent's next year's state income tax—because that is the way Californians pay for many public services provided by the State including those they vote to tax themselves for in actual state referenda. Respondents were told that the survey was being conducted for the State; our preliminary work found that most respondents accepted this way of paying for the program. Furthermore, this payment vehicle has the additional attribute of strongly invoking a budget constraint, as our design work indicated that many Californians were reluctant

---

[59] The NOAA Panel preferred this format to the open-ended alternative (Arrow, p. 4612).

See sections 4.3 and 4.5.1 of Chapter 4 and the NOAA Panel Report (Arrow, p. 4609).

to increase their income taxes.

Length of payment. A single, lump-sum payment was used in this study because the accelerated recovery program would take place in a single year and respondents pay income taxes on a yearly basis. This is a conservative strategy, as it forces respondents to confront the financial implications of their decision without having the option of paying the amount in installments over the course of several years.

For or against elicitation question. Our instrument uses a standard format for CV referendum questions where respondents are offered two voting options: *for* or *against*. Those who volunteered that they were "not sure" were recorded as such. The answers of respondents who would not vote were recorded by the interviewer (who was instructed to accept such answers as valid without further probing).

Tax amounts. We used five different initial tax amounts : $10, $25, $80, $140, and $215. The corresponding lower, second tax amounts asked of those who voted *against* at the first amount were: $5, $10, $45, $80, and $140, respectively. For those who voted *for* at the first amount, the second amounts asked about were: $25, $45, $140, $215, and $360, respectively. We chose the tax amounts to help increase the precision of the estimate of mean willingness to pay from the responses to the base survey instrument and to provide reasonable statistical power in testing whether there is a difference between the willingness to pay distributions for the base and scope versions of the survey instrument.

Description of substitutes. People typically have a range of natural resources that they can

---

Coincidentally, the survey went into the field approximately five weeks prior to the due date for California State income tax returns.

See Chapter 4 for a discussion of an split-sample test we conducted using a contingent valuation survey similar to this one to measure the effects of offering an explicit *no-vote* option.

enjoy, so that if one is not available they can use another substitute resource. To be valid, a CV scenario should describe the resource to be valued in the context of relevant substitutes. In our instrument, we presented information about a number of potential substitutes which are listed here and described in the next chapter. These substitutes and their analogue in our scenario included: (1) various other social problems that respondents might want to spend more tax money to solve (i.e, actions that are alternatives to addressing the problems caused by the DDT/PCB deposit), (2) uninjured members of the same fish species currently living elsewhere off the South Coast, and (3) uninjured members of the same fish and bird (latter in base only) species living elsewhere in California and the U.S. We also reminded respondents that members of other species of fish and birds live off the South Coast.

In the base instrument, substitutes would be needed only for an interim period until, as a result of natural processes, the four species completely recover in fifty years. In the scope version, the recovery time for the two fish was described as fifteen years. The credibility of these predictions was reinforced by mention of the recovery of other local species. The base version also described the increases in the numbers of the two bird species everywhere else in the United States. Respondents who received the base instrument were also told that consideration was being given to reclassifying the Bald Eagle and Peregrine Falcon from endangered to threatened in some parts of the country, including California. Finally, respondents were forcefully reminded just before the voting questions that some of the affected species are common elsewhere and that they all will recover on their own in fifty years.

Quantity of information. The information provided was chosen to convey the key elements of the choice (such as the information our development work showed was necessary to

See Chapter 4 and Arrow (p. 4605).

avoid possible misconceptions on the part of the respondent), and the amount presented was as much as we believed could be presented without harming the quality of the interview. In order to avoid overload and respondent fatigue, we paced the flow of information and used visual aids and questions to maintain respondent interest.

Visual displays. In-person interviews commonly use show cards to provide respondents with a visual representation of some of the material which the interviewer presents verbally. In the main study instrument, we used show cards to display lengthy lists of answer categories for closed-ended rating scale questions and to display line drawings and tables to illustrate various features of the information provided. We also used larger drawings in a separate booklet of maps and diagrams for the same purpose. In order to evaluate their ability to effectively communicate information without bias, we pretested these materials and subjected them to peer review at several points during the instrument development process.

§ 5.5 Instrument Design — Stage 2

In March, 1992, we began to test a draft instrument in one-on-one interviews using cognitive interviewing techniques (Jobe and Mingay, 1989). Our aims were to see whether the spoken text flowed smoothly when administered and whether the respondents understood the wording and the visual aids and regarded the choice they were asked to make as credible.

After further refining the survey instrument, a small number of professional interviewers

---

[60] Respondents were paid to come to an interview room provided by market research firms in various locations throughout California. We continued to conduct occasional cognitive interviews to assess new drafts at various points throughout the rest of the study, particularly between Pilots II and III.

[61] The cognitive techniques that proved to be the most useful were the retrospective think-aloud and probing techniques.

administered it under field conditions: face-to-face in the homes of respondents who were not

paid to take the interview. The survey firm contracted to conduct the surveying for this study was

Westat, Inc., a firm headquartered in Rockville, Maryland. As shown in Table 5.2, this took place

in two pretests during May and June of 1992. We debriefed the interviewers after each pretest

and revised the instrument on the basis of their comments as well as on the responses of the

pretest respondents to the survey questions.

Table 5.2 Pretests of the Base Questionnaire

Pretest	Field Period	Sample Size
1	May 16-21, 1992	57
2	June 5-14, 1992	48

The basic framework of the interview, such as the way we described the injuries and the

program to accelerate recovery, showed sufficient promise to justify moving to the pilot testing

stage.

Westat, one of the nation's largest survey research firms, has extensive experience conducting
large in-person surveys for federal and state agencies. See Appendix C-1 for a copy of Westat's
brochure.

§ 5.6 Instrument Design — Stage 3

Throughout the instrument development process, we worked to simplify the language and presentation to minimize the instrument's cognitive burden. During the stage 3 period, the survey instrument underwent many revisions as we conducted a series of pilot tests. The wording was improved. Information was added to prevent the repetition of the mis-impressions of prior respondents. The sequence of material was altered so that the material flowed naturally and held the respondent's interest.

We also performed several other important tasks during this stage: (1) we revised the instrument three times to accommodate changes in the number of the species survey respondents were asked to consider in the object of choice, (2) we conducted several split-sample tests to examine the effects of possible design features, and (3) we considered the implications of the NOAA Panel's recommendations on contingent valuation which became available after Pilot II. In particular, two issues discussed by the NOAA Panel resulted in the addition of another development stream to the questionnaire development process during this stage. This second development stream, roughly contemporaneous with Pilot III, examined interviewer influence on respondent answers and the effect of a would-not-vote option.

§ 5.6.1 Pilot Surveys

Beginning in July 1992, four pilot surveys were conducted over a period of 16 months. Pilot surveys usually differ from pretests in that they use more formal sampling techniques and larger samples. A larger number of interviewers are involved and the longer field period makes it possible to reach a greater diversity of respondents. As a result, pilot surveys provide a more detailed basis for evaluating how well the interview works in the field. Although the pilot

sampling procedures are adequate for instrument development purposes, they are less rigorous than those used in the final survey.

§ 5.6.2 Sampling and Administration for Pilots

Table 5.3 presents basic information about each pilot. The samples were designed to represent the non-institutionalized population of California age 18 years and over. Westat's trained listers canvassed and listed the dwelling units in 75 locations (segments) in ten randomly selected Primary Sampling Units (PSU's). From these listings, a specified number of dwelling units were randomly selected and fielded for each pilot. With the exception of Pilot I, the interviewers conducted a screener interview to select one respondent for the interview. The selection of the respondent was made from all individuals in the household who met the eligibility requirements: age 18 or older and owning, renting, or contributing toward the rent or mortgage of the home. In general, no attempt was made to convert refusals, and only a limited number of callbacks were made.

Table 5.3 Pilot Studies I, II, III and IV

Pilot	Field Period	Sample Size	No. of species	W-1 Design Amounts	Special Features

The final survey was based on a more rigorous design, extensive call backs, and a refusal conversion program which led to a much higher response rate. See Chapter 7 for more on sampling for the final survey.

[62] These were: San Diego, San Bernadino/Riverside, Los Angeles City, Los Angeles County, Orange, Kern, Greater Sacramento area, San Francisco Bay area, Sonoma, and Del Norte/Humboldt.

A formal screener was not used in the first pilot; rather, at each selected dwelling unit, the interviewers were instructed to use a statement provided on the cover of the survey questionnaire to identify eligible respondents.

Pilot	Field Period	Sample Size	No. of species	W-1 Design Amounts	Special Features
I	7/92	332	6	$10, $45, $80, $215	First full field test of the draft instrument.
II	8/92 to 9/92	460	6	$10, $45, $80, $215	Split-sample design comparing: a) 50 versus 150 year natural recovery period and b) two alternative placements of a single debriefing question.
III	7/93 to 8/93	324	5	$10, $45, $80, $215	Response rate test with split-samples targeted for low and high response rates.
IV	10/93 to 11/93	473	5	$10, $25, $45, $80, $215	Reversed the order in which the natural recovery and speed-up program had previously been presented, with natural recovery option presented second.

Westat conducted the data collection for each pilot using standard procedures. The interviewers attended a two-day training conducted by Westat personnel. The interviewers and the other field staff were not informed of the survey's intended use in litigation. The survey was represented as a study the State of California was conducting to learn how California citizens felt about increasing their taxes to pay for the accelerated recovery program. It was emphasized to the interviewers that there were no right or wrong answers to the voting (or any other) questions and that the goal of the study was to find out what people really felt about the topic.

For each pilot study, after the field administration of the survey was completed, Westat assembled a representative selection of about ten interviewers for a one-day debriefing session

---

There is reason to believe that response rates are higher for government sponsored surveys.

conducted by a senior Westat administrator and observed by Robert Mitchell and/or Stanley Presser. During these sessions, the interviewers were encouraged to report what sections of the survey worked well, what sections needed improvement, and any suggestions they had for wording and other types of presentational improvements.

§ 5.6.3 Pilot I

Pilot I was the first full field test of the base instrument. The debriefed interviewers reported that the scenario and the graphics held most respondents' interest, and respondents seemed to take the choice seriously. The interviewers identified problems with the wording at various points, including the first sequence of questions in Section B, which asked respondents to reveal what they had in mind about certain topics when they voted. The interviewers made numerous suggestions for improving the wording which yielded subsequent changes. In this and the following pilots and pretests, we also gained insight into how well the questionnaire was working by reviewing the answers to the open-ended questions and spontaneous comments; in both cases, interviewers were instructed to record them word-for-word as closely as possible throughout the questionnaire. Finally, the analysis of the data from the closed-ended questions was considered during revisions.

Among the changes incorporated in the Pilot II version were the addition of a prologue that, to encourage respondents to make their own judgments, presented the interview as a common practice of the State to discover public sentiment on various programs the State might conduct; a statement that the affected bird species do not migrate (which addressed the belief held by some that birds elsewhere might be helped by the program); a similar statement about the fish; information that the population of sea lions had increased greatly during the last 15 years (to

underscore the fact that they are not an endangered species); and a complete revision of the first eight questions in Section B.

In addition to implementing the wording changes listed above, we made a number of changes in the show cards. For example, we dropped as redundant a show-card we used in Pilot I that summarized information about the birds' reproductive success (Card D) and, in an attempt to improve communication in other parts of the instrument, we added show cards that listed all the species affected by the deposit (Card G1) and the reasons why the respondent might want to vote against the program (Card G2). The two maps showing the past sediment buildup and the natural recovery option were redesigned to identify the contaminated sediment layer better and to convey in a clearer fashion the progressive increase in sediment depth over the fifty-year natural recovery period.

§ 5.6.4 Pilot II

Pilot II evaluated the changes made to the instrument as a result of Pilot I and two special issues. The first concerned the length of the natural recovery period. Because there was uncertainty about the length of the recovery period, two different treatments were fielded in this pilot to explore what effects changing the recovery period would have on the scenario's plausibility. The scenario for one sub-sample used a fifty year recovery period; a comparable sub-sample used the same scenario except that the recovery period was 150 years. From the interviewer debriefings and an examination of the verbatims and responses to relevant questions, we found, overall, that respondents accepted the longer time period as credible. We concluded that we could, if necessary, lengthen the time period should such a change be needed.

Comparing the W-1 response for the 50 and 150 year treatments using a probit equation

where the slope parameter on the log of the W-1 amount is allowed to vary by treatment, we found an insignificant but suggestive differences (p=0.121), using a one-sided asymptotic t-test between the two treatments. Dropping the respondents who did not pay California taxes (an issue discussed at some length in Chapter 9), we find that the null hypothesis of no difference between the two treatments would be rejected at p=0.058. Controlling for those who thought the issue of chemical contamination in question A-1d was "extremely important" or "very important" and letting the coefficient on the A-1d dummy variable vary with the treatment allows one to reject the equivalence of the slope parameters on the log of the W1AMT at p=0.016 for the full sample and at p=0.007 after dropping the respondents who did not pay California taxes.

The second issue we evaluated with a split-sample design explored the effects of the placement of the open-ended, follow-up question asked of respondents who voted *for* the program. The question asked what the program would do that made them willing to pay for it. One sub-sample received placement 1 which put this question <u>between</u> the first voting question and the second voting question. Another sub-sample received placement 2 in which the follow-up question was asked after the respondent had answered <u>both</u> the first and the second voting questions. At issue was whether the immediate proximity of the follow-up question to the first vote question in placement 1 would reveal different insights into the respondents' valuations than placement 2. We were also interested in whether placement 1 would affect respondents' answers to the second voting question in some systematic way, such as by making them more self-conscious about their WTP responses.

A-1d is a key preference question asked in the survey before any aspect of the injuries or the accelerated recovery program is described. As a result, the response to it would not be influenced by which treatment the respondent received.

[63] See question W-6 in the main study survey instrument, Appendix A-1.

We assessed the effect of placement 1 on the answers to the open-ended, follow-up question by comparing them for the two sub-samples and found no difference. As for the effect of placement 1 on the follow-up question, we found no statistically significant differences (p=0.752) in the percentage of people who voted for or against the program in the two placements. However, when Westat debriefed the Pilot II interviewers, some reported that they found the immediate "why" follow-up question interfered with the flow of the interview: the follow-up question did not seem to follow the initial question as naturally in placement 1 as it did in placement 2.

These findings led us: (1) to ask the "why" follow-up question after the respondent had answered both WTP questions (placement 2) since the alternative offered no advantage and posed some disadvantages, and (2) to encourage interviewers in the main survey to carefully record any spontaneous comments respondents made when they answered the WTP questions.

§ 5.6.5 Pilot III

The third pilot was the first full field test after the Trustees directed that we drop the Brown Pelican from the list of affected species and value the recovery of five species (two fish, two birds, and one mammal). Other changes made to the instrument as a result of Pilot II were: 1) reworded questions A-1 and A-2 to make them easier for respondents to understand, 2) added language to clarify that the fish are only affected in one area, 3) clarification of the meaning of "sediment", 4) reassurance that the program would not stir up the existing sediment during the process of covering it, 5) clarification of the timing of the two options, and 6) a reworded B-6 to avoid having respondents think the question referred to the interviewer rather than the interview. We modified the scenario introduction further to enhance accountability by telling respondents that they would be asked later in the interview to explain why they felt the way they did about the program. We also simplified the question sequence about the Channel Islands, simplified the description of how DDT and PCB's affect wildlife and further modified questions B-1 through B-3. Finally, the show cards underwent changes designed to better communicate both the endangered and the non-endangered character of the five species.

§ 5.6.6 The Ballot-Box and No-Vote Option Study

At the same time that we were designing and implementing Pilot III, we were examining in a separate development stream whether two issues raised in the NOAA Panel Report should be implemented in the main survey we were designing for this study. One of these issues, described in Chapter 4, was whether CV surveys should offer, in addition to the *for* and *against* options in

---

Studies show (see Krosnick, 1991, for a review) that respondents are most likely to put more effort into their responses when they believe they will be held accountable for justifying their answers.

the voting questions, a would-not-vote option. The second issue was whether secret balloting should be used to avoid interviewer influence.

The results of development work on the would-not-vote option were presented in Chapter 4; the Panel's recommendation that CV surveys include this option was one of the few recommendations we did not implement. The findings of the ballot box survey we conducted to test for interviewer effects was deferred to the present chapter because of the Panel's recommendation that major CV studies should assess interviewer effects.

Because of prior commitments on the part of Westat and the need for quick resolution of this issue at a time when the instrument for this study was still in development, we chose to resolve these issues by conducting the appropriate surveys with a fully developed instrument whose design was comparable to that used in this study and that could be administered to a non-California sample by the National Opinion Research Center (NORC) of the University of Chicago. This instrument was the Exxon Valdez Oil Spill (EVOS) damage assessment survey. As described in Chapter 4, the use of the EVOS survey instrument provided an additional benefit: confirming the temporal stability of CV natural resource damage estimates obtained from this type of CV survey.

§ 5.6.6.1 Design and Implementation of the Surveys

Four new versions of the EVOS instrument were created to test how the would-not-vote and secret ballot procedures affect the WTP amounts and data quality in a CV survey that closely resembles the one used in this study.

See Carson *et al*., (1992) for a complete description of the study.

A full description of these may be found in NRDA (1994).

Version I represented the standard version of the EVOS instrument and is virtually identical to the instrument used in the original Exxon Valdez study (Carson *et al.*, 1992).

Version II, the ballot box version, was identical to the standard version except that it offered respondents the opportunity to vote in secret. After administering the elicitation question at the first (and only in this version) voting question, the interviewer was instructed to hand the respondent a paper ballot with the text of the question written on it and places for the respondent to mark his or her choices. In this version the choices were the same as Version I — for or against. After voting, the respondent was instructed to seal the ballot in an envelope provided by the interviewer, and then to place the sealed ballot in a wooden, locked, ballot box. The paper ballot was coded so it could be matched with the correct questionnaire at NORC's headquarters.

Version III, the would-not-vote version, was identical to version I except that an explicit would-not-vote option was added to the for/against categories.

Version IV, the ballot box/would-not-vote version, included both the novel features of versions II and III. Otherwise, it is identical to the standard version.

Within each treatment, four cost forms were used. As illustrated in Table 5.4, each cost form used a different set of dollar amounts as the cost of the prevention program. When the first WTP question was asked, depending upon the cost form of the questionnaire (A, B, C, or D), the respondent was told the cost was $10, $30, $60, or $120 dollars, respectively. If the respondent voted *for* the program, the second voting question was asked with a higher amount than in A-15, either $30, $60, $120, or $250, which amount depending on the cost form. If the respondent voted *against* the program, the second voting question was A-17 which incorporated a lower

---

The results from these surveys, also reported in Chapter 4, draw on comparisons between the standard version I and the ballot box and would-not-vote versions II and III. Appendix B.1 discusses the version IV results.

amount than in A-15, either $5, $10, $30, or $60, which amount depending on the cost form.

Since the two ballot-box versions did not use a second voting question, the cost forms of

questionnaire versions II and IV used only a single dollar amount and not a second amount, lower

or higher.

Table 5.4 Tax Amounts for the Voting Questions

	A	B	C	D
A-15 (Versions I, II, III, IV)	$10	$30	$60	$120
A-16 (Versions I, III)	$30	$60	$120	$250
A-17 (Versions I, III)	$5	$10	$30	$60

As noted earlier, the National Opinion Research Center (NORC) administered this survey.

None of the NORC field personnel who worked on this study were told that this study might be

used in litigation. During the field period, only a few higher-level staff in NORC's Chicago office

knew of the intended use of this study.

We conducted two pretests of selected versions. In the first pretest, 64 interviews were

conducted in the field with questionnaire versions II and IV. After modifications, version II was

pretested a second time in 26 interviews.

The field work for the main survey used a probability sample of adults chosen from 34

counties throughout the United States. This main sample was designed according to standard

---

 A report on the methods used in this study is contained in NORC (1993).

 This study was entitled the National Issues Study; and in dealing with NORC personnel, we
referred to this effort by that name.

 For a more extensive description of the sampling, see NORC (1993).

procedures although, due to time constraints, the selection of PSU's was determined by the availability of sample and sufficiently experienced field personnel. The 28 interviewers who worked on the study were trained in Arlington, Illinois on May 23-24, 1993. The interviews for the study were conducted over an eight-week period from May 26 to July 17, 1993. A total of 1182 interviews were conducted for an overall response rate of 73 percent.

§ 5.6.6.2 Pretesting for Interviewer Effects

The issue raised by the Panel was whether the presence of the interviewer in a CV survey such as the one used in this study might lead some respondents to feel pressured to vote in a socially desirable way. The Panel felt this might happen in CV surveys about natural resource damages because protecting the environment "is widely viewed as something positive" (Arrow p. 4611). In order to assess this possibility the Panel recommended that major CV studies conduct split-sample tests using a secret ballot to test for this type of interviewer effect.

We used the same split-sample methodology for this survey as described in Chapter 4 for the would-not-vote study. In what follows, we present the comparison of the standard version, questionnaire version I (N=300), with the ballot box version, questionnaire version II (N=271). A

The response rate is calculated as 1182 completions divided by 1610 eligible dwelling units. Of the 1841 households in the original sample, 159 were vacant, 56 were not dwelling units, and 16 were non-English speaking households.

On the other hand, increasing someone's taxes for any purpose would be widely viewed as something negative.

[64] Chapter 4 presents the rationale for using the Exxon Valdez survey for this test and describes the procedures used to conduct these tests.

See Appendix B.1 for a discussion of the version which contained both the ballot box and the would-not-vote options.

comparison of the answers to the first willingness-to-pay question (the ballot box procedure made it impossible to ask the follow-up, willingness-to-pay question) shows that the overall percentage voting *for* in the standard version is 52.7 percent versus 50.6 percent in the ballot box version. This difference is not statistically significant (p=0.56).

Table 5.5 compares the answers to the first willingness-to-pay question given by respondents in the base and ballot box treatments for each of the four dollar amounts used in the study. None of the four comparisons shows a statistically significant difference and, for the three higher amounts, the two versions have virtually identical percentages of respondents voting for the program to protect Prince William Sound from a future oil spill. These findings suggest that carefully designed CV surveys using a format and method of administration similar to the present survey and conducted by well-trained professional interviewers can avoid social desirability bias.

**Table 5.5 Comparison of Votes at Different Tax Amounts
for the Standard and Ballot Box Versions**

CHOICE	STANDARD	BALLOT BOX
Tax = $10	(N=87)	(N=74)
For	67.8%	56.8%
Against	32.2%	43.2%
	χ^2=2.092; p=0.148	
Tax = $30	(N=66)	(N=69)
For	56.1%	56.5%
Against	43.9%	43.5%
	χ^2=0.003; p=0.957	
Tax = $60	(N=81)	[N=65]
For	49.4%	
Against	50.6%	50.8

			%	
				49.2
			%	

$$\chi^2=0.028; p=0.868$$

Tax = $120	(N=66)		(N=	
For	33.3%	63)		
Against	66.7%		34.9	
		%		
			65.1	
		%		

$$\chi^2=0.036; p=0.849$$

We decided to use the standard version in this study because it offers important methodological advantages over the ballot box format and has no disadvantage. First, the standard version permits the use of the follow-up, willingness-to-pay questions which provide more valuation information. Second, the standard version makes it possible to follow the Panel's recommendation that, after they vote, respondents should be asked questions about why they voted the way they did. This recommendation cannot be implemented if a secret ballot is used (Arrow, pp. 4609, 4613). Third, use of a ballot box would have made it difficult to offer respondents the chance to reconsider their vote at a later point in the interview on the basis of further reflection, a factor we find to be important.

§ 5.6.7 Testing for Dichotomous Choice Response Order Effects

Prior to fielding Pilot IV, we also conducted a set of four split sample tests to explore whether the order of the dichotomous choices in the voting questions which followed the conventional practice of placing the vote *for* before the vote *against* (*i.e.,* "would you vote for the program ... or would you vote against it?") might bias responses toward voting *for*. Three out of the four tests showed that the order of the response categories did not affect how people voted. The marginally significant difference in the fourth treatment indicated that using the unconventional against–for order slightly increased the percent of people who would vote <u>for</u> the program. We concluded that continuing our use of the conventional for-against order was the conservative choice for our survey.

§ 5.6.8 Pilot IV

The survey instrument used in Pilot III was modified into the instrument used in Pilot IV. In addition to what we learned from Pilot III, the changes in Pilot IV reflected the other development efforts described above. The NORC study clarified that it would not be necessary to modify our instrument to offer would-not-vote or secret ballot options or to change the order of the voting options. Our parallel work on the scope instrument, which we discuss in the next section, required us to modify the base instrument to maintain comparability with the scope version. We also conducted several small pretests of the base instrument prior to fielding Pilot IV (see Table 5.6, pretests 1–4) which helped us to make a preliminary assessment of some of the wording changes.

Among the Pilot IV's wording changes were those influenced by development of the *scope*

See Appendix B-2 for a description of these tests.

instrument; an emphasis on the fact that only the five species are injured, and that there are no effects on human health. Greater emphasis was also placed on describing the survey as part of an ongoing effort on the part of the State to learn what people think about new programs. In the scope instrument, this was needed to help legitimate for some respondents why the State was concerned about two fish species.

According to the Pilot III interviewers, some respondents complained that the reasons listed just prior to the voting questions as to why they might want to vote for or against the program seemed out of balance because only one reason *for* was given versus four reasons "against". In order to modify this appearance of imbalance, we reworded the reasons "for" and the shift to the reasons "against," and we reduced the number of reasons to vote against to three in the Pilot IV instrument by integrating the first two into a single reason.

The show cards for Pilot IV used a new drawing of the falcon designed to better differentiate it from the pigeon some interviewers felt it resembled, and we connected the pictures to the names of the species on the card that showed their endangered status (Card F).

We adjusted the order in which some material was presented to prevent the possibility of a non-conservative bias. Specifically, we reversed the order in which we presented the two action options: in order to be conservative, we presented the accelerated recovery program first and the

---

Recent research indicates that when a researcher gives people information about two options and asks people to choose between them, the order in which the information is presented may, in some instances, influence choices (Krosnick, Li, and Lehman, 1990). For example, in typical, everyday conversations, people tend to provide less important, background information first, and the more important, foreground information second. Respondents might presume that the interviewer is following such conversational conventions and is providing the information he or she believes is more important after providing the less important information. To the extent that this occurs in this survey, it would lead respondents to make decisions placing greater weight or importance on the information provided later. Thus, respondents might be biased toward endorsing whichever option was described last.

let-nature-take-its-course option last.

We also adjusted the order in which we presented some rating scale categories. With regard to rating scale questions, people are inclined to select alternatives presented early over those presented later (Carp, 1974). Consequently, as a conservative measure, we reordered all of our visually presented rating scale items by listing first the response alternatives that expressed negative attitudes toward programs designed to protect the environment (*e.g.*, questions A-1 and A-2). A fifth design point, $25, was also introduced between $10 and $45. This was done to help determine whether $25 or $45 was likely to be more useful in comparing responses from the base and scope versions of the survey.

§ 5.6.9 Additional Base Pretests

We conducted four small pretests of base versions during the six months prior to fielding the final study. Because we were also conducting pretests of the scope version at the same time, the base pretests were not consecutive, as shown in Table 5.6. Pretests 2 and 4 were conducted prior to Pilot IV, and helped with the design of that pilot. Pretests 7 and 8 occurred later in the design process and pretested, respectively, changes we needed to make in base to ensure comparability with scope and changes associated with the reduction in the number of species from five to four.

The sampling frame for the base pretests consisted of between 6 and 8 PSU's. Because these pretests were to give us quick field tests of our ongoing revisions, no attempt was made to convert refusals, and the interviewers made few, if any, call backs. Table 5.6 describes each of these pretests and the role it played in our instrument development for this study.

Table 5.6 Additional Base Instrument Pretests

Pretest	Field Period (1993-94)	Sample Size	Role of the Pretest
2	8/14 to 8/22	44	After Pilot III, we changed the instrument in a number of ways, including the sequence of the programs. This was the first of two pretests we conducted to see if further changes were needed before conducting Pilot IV.
4	9/4 to 9/15	57	Second Pilot IV pretest. Among other things, it checked the feasibility of asking a follow-up revision question in the W sequence of questions (Question W-7).
7	11/17 to 11/23	49	Introduction of new language, from a scope draft, intended to better communicate where the species are harmed, that humans are not harmed, and the nature of the substitutes.
8	1/13 to 1/27	116	First use of instrument after dropping sea lions from the set of species affected by the deposit. Also tested revisions in graphics such as Card F.

§ 5.7 Development of the Scope Instrument

We followed the NOAA Panel's recommendation to examine responsiveness to scope by conducting a split-sample test in the main study. One sub-sample's choices for the injuries described in the base instrument (two fish and two bird species with a 50 year natural recovery period) was compared with another sub-sample's choices for the reduced set of injuries (two fish species with a 15 year natural recovery period) presented in a scope instrument.

There are at least three reasons why we compared the base injury scenario with a smaller rather than a larger injury scenario in this study. First, it appeared that evaluating a larger set of injuries would be more complicated and would require more time and higher cost. Second, because the degree to which the injury can be reduced is limited by no injuries, a smaller injury

scenario is likely to be a more credible demonstration of sensitivity to scope. Finally, a larger injury would have raised an ethical issue. Respondents tend to regard information they receive during a government-sponsored interview as authoritative. Describing a larger injury would have run the risk of unnecessarily alarming citizens about the state of the environment.

The reduced set of injuries we decided upon for the scope version was the two fish component of the base set of injuries with a natural recovery period of fifteen years. Alternative configurations of species would have faced plausibility problems. Because some respondents were aware of the two birds' place in the local ecosystem, it would have been hard to include just one of them in the reduced injuries. A single fish species injury would have seemed implausible to some respondents.

In this section we describe how we modified the base instrument to create the scope instrument. The primary modification involved the section of the questionnaire that described the injuries. It was also necessary to modify some other sections because we found in the focus groups and pretests that some respondents considered the reduced set of injuries improbably small for the State to be concerned about them. To the extent that respondents held this view, they tended to imagine that the program might do more than just speed up the recovery of the two fish species. As noted above, the changes we made to develop a plausible choice in the scope version also had to be made in the base instrument so that the two would be comparable in every respect except for the size of the injury.

§ 5.7.1 Scope Focus Groups

We began the process of designing the scope instrument in June 1993. During June and July of that year, we conducted five focus groups to explore how we could adapt the base instrument to present a set of injuries consisting of reproduction problems for the White Croaker and Kelp Bass which would disappear in 15 years without the accelerated recovery program. (For convenience, we will refer to these as the scope focus groups.) Because the two bird species were not included, the injuries were described as occurring only in the immediate area of the deposit (marked in red on Map 3 in Appendix A-2) off the Palos Verdes Peninsula. A few months later we conducted four additional focus groups whose participants came from two different demographic groups. Table 5.7 gives the dates and locations of the nine scope focus groups.

Table 5.7 Scope Focus Groups

Focus Group	Location in CA	Date Conducted
1	Santa Monica	June 17, 1993
2	Orange	June 18, 1993
3	San Francisco	June 30, 1993
4	Sacramento	July 1, 1993
5	San Diego	July 2, 1993
6 & 7	San Francisco	October 28, 1993
8 & 9	Los Angeles	October 29-30, 1993

The participants in scope focus groups 1–5, which were held in different parts of California, shared the same mix of personal characteristics as the participants in the base focus groups. Because we already had a base instrument to build on, we bypassed one-on-one

interviews and immediately began group interviews. We first administered a draft scope instrument (through the willingness-to-pay section) verbally to the group as a whole, with the participants recording their answers privately in a response booklet. The WTP dollar amount was varied from group to group. At the end of the group interview, the moderator led a discussion to ascertain the participants' reactions to the scope version and to learn what they felt about particular issues.

Many of the focus group participants did not think reproduction problems in two fish species were worth paying for as long as human health was not affected and the fish were not endangered and would recover anyway in 15 years. Some questioned why the State would conduct a survey about this small an injury. We addressed this by modifying the way we portrayed the circumstances of the survey to imply that the State did surveys like this on possible new programs and this happened to be the program this interview was about. Further, some scope participants focused on the possible human health implications of eating the fish. The level of this concern appeared to be due to the fact that the scope injuries were restricted to fish. In order to minimize this concern, we enhanced the assurances given in the instrument that the fish injuries did not pose a threat to human health.

In October, 1993, we conducted four additional focus groups with homogeneous participants to understand how particular populations react to the scope injury. We recruited women for two groups in San Francisco and minorities for two groups in Los Angeles. Because they had a larger number of participants who voted *for* the scope program, the two Los Angeles groups provided some useful insights about why low income people chose to vote for the program. These participants were particularly concerned about the health implications of eating

----

The San Francisco groups were less informative about why people favored the program to

fish contaminated by the chemicals and distrustful of the government, which led some to believe the injuries were likely to be greater than described. We further revised the instrument to provide stronger, more plausible assurances that human health was not threatened by the situation and that the injuries were limited to those described.

§ 5.7.2 Scope Pretests

Following the first scope focus groups, we conducted a series of four pretests to evaluate various versions of the draft scope instrument as it evolved. Table 5.8 describes each pretest

Table 5.8 Scope Pretests

Pretest	Field Period (1993-94)	Sample Size	Role of the Pretest
1	8/8 to 8/14	44	First field test of the scope instrument.
3	8/25 to 8/30	54	Second test after the first round of revisions. Those who mentioned health in the verbatims tended to be more likely to vote for the program than those who did not.
5	9/18 to 9/25	40	Revisions tested in this pretest focused on improving the plausibility of asking respondents to value speeding up the recovery of two fish species in one local area. They also included a new question (W-7) which offered the respondents the opportunity to change their vote if human health was definitely not affected in the situation.
6	11/17 to 11/23	44	In pretest 3, we continued to find some respondents found it difficult to believe the injuries were restricted to the two fish. After conducting the last round of scope focus groups, we used this pretest to test further revisions intended to better communicate that only the two fish species were affected.

prevent the reduced injuries because only two people in each group said they would vote for the program.

and the role it played. The pretests took place during a six month period beginning while Pilot III (base instrument) was in the field and were interspersed with the base pretests described earlier. The sampling frame for the scope pretests was the same one used for the earlier base pretests and, as was the case with the base pretests, no attempt was made to convert refusals and the interviewers made few, if any, callbacks.

§ 5.8 Spanish Translation

In California, over 31 percent of people 18 years and over speak a language other than English at home; and, of those who speak a language other than English at home, a large majority speak Spanish. Unfortunately, the available census data did not provide information that would have allowed us to ascertain what fraction of California Spanish–speaking households had someone meeting our sampling criteria who is a fluent English speaker. In the absence of this information and given the relatively large size of the Spanish-speaking population in California, in November of 1993, NRDA retained Aguirre International (hereafter, Aguirre) to prepare Spanish versions of the main study survey instruments. At this stage in our instrument development, we felt that the English version of the survey instrument — that used in Pilot IV— was far enough

---

Language Use and English Ability, Persons 18 and Over, by State, in Education and Language Data for States: 1990 Census, December 15, 1992.

The next two most common non-English languages are Chinese (6.7%) and Tagalog (5.4%), the latter spoken by Filipinos (*Languages Spoken at Home by Persons 5 Years and Above, by State: 1990 Census*, in Education and Language Data for States: 1990 Census, December 15, 1992).

Aguirre has considerable experience developing effective Spanish translations of survey instruments and advising clients on how to implement surveys among Hispanic and other ethnic groups.

along to warrant the initiation of this effort.

§ 5.8.1 Survey Development

Aguirre's task was to develop a Spanish translation that not only met the five objectives of the development process outlined above but also accommodated the idiosyncrasies of California speech patterns and was sensitive to California's Hispanic cultures. The translation had to be understandable to California's Hispanic populations (*e.g.*, Mexicans, Cubans, Nicaraguans) who speak slightly different Spanish dialects, and yet not offend the sensitivities of respondents who may be Spanish language purists. Aguirre's translation also had to be sensitive to cultural differences among California's hispanic populations. For example, those schooled in the U.S. would be accustomed to the presentational format used in the survey instrument (*e.g.*, answer categories, diagrams, and maps), whereas those schooled in Spanish-speaking countries may not be. At each stage in this development process, NRDA worked closely with Aguirre to ensure that the five objectives of the development process were not compromised and to ensure that the essential elements of the choice were not lost in the translation.

The first step in Aguirre's development work was to translate the Pilot IV version of the survey and then to make an independent, reverse translation of this Spanish version back into English. Aguirre conducted group discussions and informal, cognitive interviews to aid in the translation. After the instrument was revised to address inconsistencies between the forward and back translations, Aguirre conducted additional cognitive interviews in various locations that were selected to represent the regional differences in California's immigrant populations. Throughout this second step, the draft translation of the base instrument was continually revised and

improved. Aguirre was also asked at this point to translate into Spanish and back into English those sections of the scope version of the questionnaire that differed from the base version and to translate the Pilot IV screener.

During Aguirre's field work, Trustee representatives directed NRDA to omit one of the species from the set of injuries to be valued. This revision so late in the survey development slowed down our work on the translation so that Aguirre had less time than planned to incorporate feedback from their interviewers on their field experiences for pretesting the Spanish base and scope instruments in the field.

Westat's bilingual interviewers conducted the first full field pretest of the Spanish base and scope instruments in late January. Eighteen interviews using the base instrument were collected from January 16 through January 20. Twenty-one interviews using the scope instrument were collected from January 23 through January 27. In most respects the Spanish translation performed well, given the linguistic and cultural complexities of rendering the choice elements into Spanish; but the debriefing showed that additional development work would be needed before the Spanish instruments could be fielded in the main survey. Because of the imminent deadline for delivery of the main study questionnaire, we decided that Spanish versions of comparable quality to the English versions of base and scope could not be readied in time for the field deadline.

As a result of being unable to field Spanish-language versions of the main survey instruments, we treated the Spanish-language speakers like other non-English speakers, *i.e.*, as

---

In the course of their instrument development work, Aguirre identified an additional complication in conducting this interview in Spanish. Some Spanish-speaking Americans who received their high school education in English and who would choose, because of cultural pride, to be interviewed in Spanish, do not have an adequate command of Spanish for this purpose. Aguirre believed these respondents would be fully capable of taking the interview in English.

ineligible for the survey. The issue of survey eligibility is discussed in Chapter 7.

§ 5.9 **Final Pretesting**

The last pretests for this study were conducted in January and February, 1994. In the case of both the base and scope instruments, a "pre-main" pretest was conducted, after which a few minor changes were made to improve comprehension, and the revised version received a final pretest before delivering the finished instrument, without further change, to Westat for use in the main survey.

§ 6 Structure of the Main Study Questionnaires

§ 6.1 Introduction

This chapter describes section-by-section the wording, format, and presentation in the base and scope questionnaires used in the Lost Use Value study. All quoted text in this chapter is common to both the base and scope instruments unless otherwise indicated. Both complete survey instruments, including the show-cards and a reproduction of the map-diagram booklet, are provided in Appendices A.1 and A.2.

To avoid self-selection bias from people deciding to be interviewed because of their interest in the specific subject matter of the survey, prospective respondents were told that the State of California was conducting the study to get their "opinion on issues that may concern you such as education, the environment, and crime" (Westat, 1994a). If potential respondents asked for more information about why the survey was being conducted or what it was about, the interviewers were instructed to use only the replies provided on a laminated Q and A card. For example, if asked "Why are you doing this survey?" they were to say: "The study will provide information so State policy makers can understand how people like yourself feel about these issues." If asked a question like, "What is this survey about?", they were to say:

> We are faced with many problems in California today. This study is about some of these problems and issues. Some may be of concern to you, others may not. The study attempts to find out how Californians feel about some of the problems facing the state today.

Some questions and answers (Q & As) to questions we anticipated would be most frequently asked were included in the questionnaire text. The Q & A card (reproduced in Appendix A.3) contained additional Q & As.

This typeface will identify lengthy direct quotations from the language of the questionnaire or interviewer's manual.

§ 6.2 Section A — Introductory Questions

The first set of questions (A-1A to A-1F) in the interview proper asked how important six state-wide issues were to the respondent personally.

A-1. Let's start by talking for a moment about some issues in California. Some may <u>not</u> be important to you, others may be. First, (READ X'D ITEM), is this issue

> SHOW CARD A

not important at all to <u>you personally</u>, not too important, somewhat important, very important, or extremely important? (READ EACH ITEM, BEGINNING WITH X'd ITEM; CIRCLE ONE CODE FOR EACH; RE-READ STEM AS NECESSARY.)

This, and the following series of problems (A-2 described below), encouraged the respondent to think about a broad range of current policy issues as a reminder that speeding up the recovery of the affected species is just one of many public goods. Two — "reducing crime" and "improving education" — reminded the respondent of issues that are of great concern to some Californians at the present time. "Finding ways to reduce state taxes" was chosen to remind the respondent of the linkage between state programs and the level of state taxes. "Maintaining library services" represented one of a range of local community spending issues. "Reducing air pollution in the

---

Any questionnaire text in capital letters is an interviewer instruction and is not read to the respondent.

These instructions cue the interviewer to show Card A. This card lists five answer categories from "not important at all" to "extremely important". See Appendix A.1.

Following standard survey practice to minimize response order effects, the order in which the six items were asked was randomized. The interviewer was instructed to begin with the item marked "X". Each item had an approximately equal chance of being asked first.

Highlighting taxes helps address the NOAA Panel's recommendation to deflect "warm glow" motivations (Arrow *et al.*, 1993; p. 4609).

cities" was an environmental issue not directly related to coastal natural resources while

"protecting coastal areas from oil spills" was directly related.

Question A-2 shifted the respondent's attention to the fact that the state spends money on various programs, both non-environmental and environmental. The respondent was asked to say whether he or she wanted the state to reduce, increase, or have the amount of money the State was spending on these programs "stay the same". One environmental program directly related to the injuries, "protecting endangered wildlife species", was included in this series.**A-2. The State of California spends money on many programs for many different purposes. I'm going to read a list of some of these programs. For each one, I'd like you to tell me whether you think the money the State is spending on these programs should be**

> **SHOW CARD B**

reduced a great deal, reduced somewhat, stay the same, increased somewhat or increased a great deal. First, (ITEM)? (READ EACH ITEM, BEGINNING WITH X'd ITEM; CIRCLE ONE CODE FOR EACH; REPEAT ANSWER CATEGORIES, AS NECESSARY.)

§ 6.3 Section A — Description of the Elements of Choice

The presentation of the elements of choice, which began at this point, provided the

circumstances of the choice relevant to the decision the respondent would later be asked to make

— to vote *for* or *against* the accelerated recovery program (*i.e.,* the object of choice). Among

the material included was a detailed description of the injuries, their cause, how long it will take

the affected species to recover with and without the accelerated recovery program, and how the

program would work.

The interviewer training for this study emphasized reading this material in a way that

---

The other programs involved: new state prisons, public transportation in Los Angeles, pay raises for professors at state universities, homeless shelters, and lifeguards at state beaches.

This card lists five answer categories from "reduced a great deal" to "increased a great deal".

would maintain respondent interest and enhance comprehension. The interviewer's manual

summarized this emphasis:

> This survey differs from most of the surveys you may have conducted because a central portion of the questionnaire has you read a narrative to the respondent. The narrative material is illustrated by maps and show cards which you show the respondent. Reading this type of material requires a somewhat different approach than reading regular question material: in effect, in presenting the material, you have to tell a story. Throughout our earlier pretests and pilot studies, we have found that the text goes smoothly and that most respondents find the material interesting.
>
> The narrative material is intended to provide the respondents with information about the situation on which they are asked to vote in questions W-1 through W-3. **It is crucial that the respondent listen carefully to what you are reading so that he/she can make an informed decision when responding to these questions.** Because of the volume of material you will be reading, there is a risk that some respondents will become bored or disinterested at some point during the interview. You should do your utmost to keep the respondent's attention throughout. You will find the maps and show cards and pointing out certain features in them helpful in this regard. Of particular importance is how you present the material. It should be read in a manner that is conversational and interesting. To do this, you need to make use of effective "body language" and use a tone of voice and manner that is interesting to listen to. Avoid reading in a monotone or conveying the impression that you are bored. [Westat, 1994a; p. 4.6]

At places noted in the text the interviewers showed the respondents visual aids. These

materials were designed and pretested to help the respondents visualize important aspects of the

scenario and to help them understand the material that was being read to them. The visual aids

consisted of two sets of booklets that were spiral bound for ease of use by the interviewers. For

the base survey set, one booklet, measuring 11 ½" x 14", contained six color visual aids; and the

second, measuring 8 ½" x 11", contained fifteen show cards printed in black and white on light

cardboard stock. When administering the scope version, the interviewers used a separate but

---

See Appendix A.1.

  These consisted of three maps and three diagrams; for interviewer convenience, all the maps and diagrams were referred to as *maps*.

comparable set of visual aids modified to fit that scenario.

Turning now back to the survey text, after question A-2, a transition was made to introduce the respondent to the subject matter of the survey. This was done in two steps. The first introduced a credible rationale for why the respondent will be asked whether he or she would vote to tax his/her household for a program such as the one presented in the survey. The interviewer says:

> These are just a <u>few</u> of the things the State of California spends tax money on. Proposals are sometimes made to the State for new programs. The State does not want to undertake new programs unless taxpayers are willing to pay for them. <u>One</u> way for the State to find out about this is to give people like you information about a program so that you can make up your <u>own</u> mind about it.

In order to help avoid creating the impression that there was a preferred response to the choice questions, the respondent was told that people responding to this type of interview had different views about the program. Specifically,

> In interviews of this kind, <u>some</u> people think the program they are asked about is <u>not</u> needed;   others think it is. We want to get the opinions of <u>both</u> kinds of people.

At this point in the interview a question was asked to involve the respondent:

To help the interviewers avoid inadvertently using the wrong set of visual materials, all base materials had blue covers and all scope materials had yellow covers.

The textual material in the questionnaire (see Appendix A.1) is often presented in very short paragraphs to help the interviewers keep their place. This convention has not been maintained in this chapter to save space.

Words are underlined throughout the interview text at places where the interviewers were instructed to emphasize words to help convey the passage's meaning and to hold the respondent's interest by making the narrative interesting.

A stop sign symbol was an instruction to the interviewers to pause before continuing.

A-3. Have you ever been interviewed like this before to get your opinion about whether the State should or should not spend tax money for a particular purpose?

The second step in the transition introduced the specific program the respondent was asked about later in the interview. Wording was used that emphasized the routine nature of this type of inquiry:

In the past, people have been asked about various types of programs. In this interview, the particular program I am going to ask you about involves two types of ocean fish and two types of birds [*two types of ocean fish*]. These fish and birds [*These fish*] are producing fewer young than normal in one particular area.

Respondents were next given an overview of what to expect in the interview. In order to encourage thoughtful consideration of their decision, they were told in advance that they would be asked to explain their choice.

First, I will tell you about what is happening to them. Then, I will tell you about the cause. Then, I'll ask you whether or not you think anything should be done about this. I will also ask you to tell me why you feel the way you do.

The interviewer next showed the respondent two maps: Map 1 located the South Coast in relation to the rest of California, and Map 2 showed the South Coast in more detail, including the location of the former DDT plant and the deposit on the ocean bottom off the Palos Verdes Peninsula.

```
┌─────────────────────┐
│  SHOW MAP 1         │
└─────────────────────┘
```

Here is a map of California. The situation I am going to tell you about is located

---

Variants in wording in the scope version will be presented in italics and brackets. Longer differences in wording will be noted in the text.

This technique of inducing accountability at the start of an interview has been shown to promote optimal respondent effort. See Tetlock (1983).

along this <u>one</u> part of the California coast, the <u>South Coast</u> ⬆ .

[**base only**] This area includes the ocean here ⬆ , the shore here (⬆ TRACE

SHORELINE) and also these islands ⬆ , the Channel Islands.

STOP UNTIL R IS FINISHED LOOKING AT MAP 1

SHOW MAP 2

In order to provide a break in the narrative, respondents who were not interviewed in Los

Angeles or Orange County were asked A-4, and those interviewed in Los Angeles or Orange

County, A-5.

A-4. Have you ever lived in Los Angeles County or Orange County?
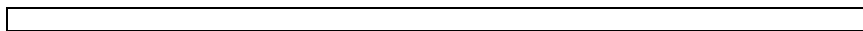
A-5. How many years have you lived in <u>this</u> county?

After identifying the geographical area of concern, the text described the affected species.

A-6. Many species of fish and birds [*of fish*] live off the <u>South Coast</u>. Four [*Two*]
of these species are having problems <u>producing</u> young [*<u>producing</u> young in
<u>one</u> place off the South Coast*].

[**base only**] I'll describe these reproduction problems beginning with the <u>fish</u>.
Two species of fish are having problems <u>producing</u> young in <u>one</u> place off the
South Coast.

[**both base and scope**] These are <u>White Croaker</u> and <u>Kelp Bass</u>. This card shows
what these fish look like.

This upward arrow symbol is an instruction to the interviewer to point to the relevant feature on
the map or card.

This card showed black and white line drawings of each fish. This format was chosen as a
conservative way to inform respondents about what each species looks like.

SHOW CARD C

(STOP) UNTIL R IS FINISHED LOOKING AT CARD

The Kelp Bass is sometimes called <u>Calico Bass</u>. Unlike <u>some</u> species of fish, these two do <u>not</u> travel up and down the coast but generally stay in one place where they live and breed.

SHOW MAP 2 AGAIN

Please look at the place marked in red on the map. ⬆ It is near Los Angeles harbor between Santa Monica Bay and San Pedro Bay. This is the place where scientists have found that the White Croaker and Kelp Bass produce <u>fewer</u> young than elsewhere. However, as millions of these two fish live in <u>other</u> places along the California coast, neither species is in any danger of becoming extinct.

The information about how many other White Croaker and Kelp Bass live along the California coast informed respondents about an important "undamaged substitute commodity". The next portion of the instrument, which described the injuries to the two birds species, appeared <u>only</u> in the base instrument. It described the nature of their reproductive problems and how these problems were limited to just the species of these birds located in the South Coast.

Two of the many species of birds living along the South Coast <u>also</u> have reproduction problems.

SHOW CARD D

They are <u>Bald Eagles</u> and <u>Peregrine Falcons</u>. These <u>eagles</u> and <u>falcons</u> along the South Coast tend to stay there all year long. Back in the 1940s, about 24 pairs of Bald Eagles and 20 pairs of Peregrine Falcons were successfully (STOP) hatching their eggs in the South Coast.

By the 1950s, the eagles and falcons in this area were having trouble producing

_____

See NOAA Panel recommendation (Arrow, pp. 4608-4609).

This card contains line drawings of the Bald Eagle and Peregrine Falcon.

young, mostly because their egg shells were too thin and the chicks did not hatch. As a result, the local populations of Bald Eagles and Peregrine Falcons disappeared from the South Coast.

About ten years ago, scientists began bringing adult falcons and eagles from outside the South Coast and releasing them on some of the Channel Islands. The scientists hoped these birds would be able to reproduce naturally and re-establish themselves in the area. Thus far, however, these birds have usually not been able to hatch any of their eggs.

```
┌─────────────────────────────────┐
│  SHOW MAP 2 AGAIN               │
└─────────────────────────────────┘
```

Unlike the White Croaker and the Kelp Bass, which only have problems in this place (⬆ TO PLACE MARKED IN RED ON MAP 2), these birds are having reproduction problems everywhere they live along the South Coast (⬆ TRACE SHORELINE), including the Channel Islands ⬆ .

In accordance with the NOAA Panel's recommendation (Arrow, 1993; p. 4609), several different checks on respondent understanding and acceptance of the scenario were used in this survey. One type of check was a question like the following which gave the respondent the chance to clarify any part of the injury description by having it repeated:

[both base and scope] A-7. Is there anything I have told you about these four fish and bird [two fish] species that you would like me to repeat?

The answer categories to A-7 were "yes" or "no". Those who answered "no" were skipped to section A-8. Those who said there is something they would like to have repeated are asked an open-ended question:

A-7A.   What is that?

This is the first of several questions in the survey which required the interviewers to record the

_____

See survey instruments in Appendices A.1 and A.2 for skip patterns.

words used by the respondent in answering the question. The interviewers were instructed to record on the questionnaire what the respondent said as closely as possible, asking the respondent to pause, if necessary, so an answer or comment could be completely transcribed. The importance of accurately recording the comments in this interview, both the answers given in response to specific questions like this and remarks made by the respondent at any other place during the interview, was emphasized in the training and in the interviewer's manual (Westat, 1994a; pp. 4-15). During training the interviewers practiced recording verbatims. For recording the verbatims, as for recording the responses to all questions, the interviewers were instructed to use a ball point pen.

The interviewers were instructed to use nondirective probing techniques to clarify respondent answers to open-ended questions when the answers were vague or did not adequately answer the question. Such probing is a standard survey procedure used to refocus respondent's attention on the question. It requires the interviewer to find a way to get the respondent to elaborate or think about an incomplete or irrelevant answer without influencing the content of an answer. The interviewers were restricted to using only probes similar to those on the list of standard probes or probes specified for particular questions.

The material that followed immediately after A-7A appeared only in the base instrument, as it described the endangered status of the two birds. (The fact that the two fish species were not endangered had already been made clear in both instruments.) This material described the

---

The interviewers who conducted this study were already familiar with verbatim recording as a result of their general training as Westat interviewers.

Chapter 5 of the interviewer's manual (Westat, 1994a) is devoted to probing.

current status of these birds in California and other states, and the population increases both birds

have been experiencing everywhere else but in the South Coast. This material and the information

that the birds were being considered for reclassification from endangered to threatened provided

respondents with information about undamaged substitutes (Arrow, pp. 4608-4609) for the South

Coast Bald Eagles and Peregrine Falcons. We used two show cards, E and F, to help convey this

information.

```
┌──────────────────────────────────┐
│                                  │
│  SHOW CARD E   (STOP)            │
│                                  │
└──────────────────────────────────┘
```

As you can see on this card, the two species of fish are <u>not</u> in danger of becoming extinct and are therefore <u>not</u> listed as endangered. However, the eagles and falcons I told you about <u>are</u> listed as endangered by the State of California. <u>At present</u>, these birds are <u>also</u> listed as endangered in most of the other states where they live.

```
┌──────────────────────────────────┐
│                                  │
│  SHOW CARD F                     │
│                                  │
└──────────────────────────────────┘
```

This card compares how these birds are doing in the rest of California and the rest of the United States ⬆ with how they are doing along the South Coast ⬆.

```
┌────────────────────────────────────────────────────────┐
│                                                        │
│  (STOP)   UNTIL R IS FINISHED LOOKING AT CARD F        │
│                                                        │
└────────────────────────────────────────────────────────┘
```

As you can see, in the <u>rest</u> of California and in the <u>rest</u> of the United States, these birds are <u>increasing</u> ⬆ in number. For example, at present, most of the 100 or so pairs of Bald Eagles that live in <u>other</u> parts of California ⬆ are <u>successfully</u> hatching young. Because the eagles and falcons are <u>increasing</u> in these areas ⬆, consideration is being given to <u>reclassifying</u> them from <u>endangered</u> to <u>threatened</u> in some parts of the country, including California.

Along the <u>South Coast</u>, however, the eagles and falcons are <u>not</u> increasing ⬆. This is because <u>no</u> eagles have hatched young on their own and only <u>rarely</u> have <u>some</u> Peregrine Falcons been able to do so. (STOP)

_____

As of the time the main study field effort began (March 8, 1994).

At this point the material is again common to both base and scope and a transition is made from a description of the injuries to their cause.

Many scientists have studied <u>why</u> these four species of fish and birds [*two fish species*] are having reproduction problems along the <u>South</u> <u>Coast</u> but <u>not</u> elsewhere along the California coast.

Some of these scientists work for the Federal Government, others work for the State, and others are independent researchers at California universities. They agree that these reproduction problems are caused by a deposit of <u>two</u> chemicals that are trapped in the sediment on the bottom of the ocean. These chemicals are <u>DDT</u> and <u>PCBs</u>.

Next, two prior knowledge questions were asked, and a definition of DDT and PCB's and their common uses were provided.A-9. Before today, had you heard anything at all about DDT?

A-10. How about PCBs? Had you heard anything about them before today?

(As you may know,) DDT is a pesticide that was developed during World War II. It was found to be a cheap and effective way to kill insects like mosquitos. PCBs are chemicals that were developed around the same time and were used in electrical transformers and for other industrial purposes.

The next portion of the narrative described how the deposit of DDT and PCB's was formed. Map 3 included an inset which enlarged the deposit area and showed the location of the sewage treatment plant and the DDT plant. The fact that the DDT plant went bankrupt and was torn down is emphasized to avoid having respondents protest that it is not fair for them to pay for the injuries because the company is responsible.

The NOAA Panel recommended that the survey be designed to deflect the "dislike of big business" (Arrow, p. 4609).

SHOW MAP 3

This big circle (⬆ TO BIG CIRCLE) is a blow-up of this small circle (⬆ TO SMALL CIRCLE.) The place marked in grey (⬆ TO GREY AREA IN BIG CIRCLE) shows the location of the deposit of DDT and PCBs on the ocean floor that causes the problems I have described. This deposit (⬆ TO GREY AREA) is about five miles [STOP] long and two miles wide. [STOP]

The biggest source of these chemicals was a factory, located here ⬆, which was at one time the world's largest producer of DDT. Over a period of thirty years, beginning in the late 1940s, this factory sent tons of waste DDT into the Los Angeles County sewer system where it went to this sewage treatment plant ⬆ and was released with other treated wastes into the ocean through these underwater sewer pipes. ⬆

A smaller amount of waste PCBs from other sources also went out the sewer pipes in the same way. [STOP] Back in the 1940s, 50s, and into the 1960s, there was little recognition that DDT and PCBs could affect fish and wildlife [fish]. When this became clear in the 1970s, sending these two chemicals into the ocean through the sewers was stopped. The federal government also severely restricted the use of both DDT and PCBs. As a result, the DDT factory in Los Angeles went bankrupt [STOP] and was torn down. [STOP]

SHOW MAP 3 AGAIN

Even though no new DDT or PCBs have been put into the sewers for about 15 years, the old DDT and PCBs located in the grey area (⬆ TO GREY AREA) have continued to affect the four species of fish and birds [the two fish species in the one place I told you about].
Here's how this happens. Because these two chemicals do not dissolve in water, they gradually fell to the ocean bottom. Once they reached the bottom, they remained there trapped as part of the sediment.

A portrayal of the deposit's stability over time was included in the scenario because our pretesting found some respondents imagined that ocean currents or earthquakes might disperse the contaminated sediment over a larger area.

This sediment — made up of things like sand and dirt — is very stable. It lies more than a mile offshore under water more than 100 feet deep where there are no strong ocean currents. Therefore, the contaminated sediment has

remained in this location ⬆, for over 15 years, where it is slowly being covered by new, uncontaminated sediment.

The next question asked the respondent if he or she had prior personal knowledge of this deposit. We used a follow-up question, A-11A, to obtain information to check if the respondent had this particular deposit in mind.

A-11. Before today, had you heard anything about the DDT and PCBs that are located in this particular place? (⬆ TO GREY AREA)

A-11A. What have you heard? [OPEN-ENDED]

The next portion of the scenario described how the deposit of DDT and PCBs caused the injuries.

A-12. The federal, state, and university scientists I mentioned earlier have conducted studies of the effect of this deposit ⬆. They know that DDT and PCBs can build up in the bodies of some fish and birds [*some fish*] when the food they eat has these chemicals in it. According to the scientists, the only animals [*fish*] that are affected by this deposit are the four [*two*] species I told you about. This is because they all feed [*they live and feed*] in this particular place.

The following material described how DDT and PCB's move through the food chain. In the base instrument, Card G illustrated this process. In the scope instrument, a parallel card (Card D) which showed only the lower portion of the base instrument's Card G (as only this portion was relevant to the two-fish scenario) was used. To accommodate the different diagrams, each version used somewhat different wording to describe how the affected species absorb DDT and PCBs into their bodies. In the base instrument, the wording was as follows:

SHOW CARD G

This drawing shows how this happens.

133

STOP UNTIL R HAS HAD A CHANCE TO LOOK AT CARD G

These ⬆ are small animals that live in the sediment on the ocean bottom. When they get food from contaminated sediment, they absorb DDT and PCBs into their bodies. When they are eaten by other <u>larger</u> animals, like this fish which is feeding on the bottom ⬆, the DDT and PCBs can be absorbed into the body fat of the larger animals.

(As you know,) This also happens when <u>larger</u> fish eat the <u>smaller</u> fish ⬆, when birds like this ⬆ eat <u>contaminated</u> fish, or when birds like this ⬆ eat other birds that have eaten contaminated fish.

REMOVE CARD G

Although the amount of DDT and PCBs in the bodies of the four species is <u>high</u> enough to affect their ability to reproduce, the amount is <u>not</u> enough to affect the adult fish or birds in any <u>other</u> way.

In the scope instrument, the wording was as follows:

SHOW CARD D

This drawing shows how this happens.

STOP UNTIL R HAS HAD A CHANCE TO LOOK AT CARD D

These are small animals that live in the sediment on the ocean bottom. When they get food from contaminated sediment, they absorb DDT and PCBs into their bodies. When they are eaten by the White Croaker and Kelp Bass, the DDT and PCBs are absorbed into their body fat. When the fish have a high enough level of DDT and PCBs in their bodies, their ability to reproduce is affected.

REMOVE CARD D

Although the amount of DDT and PCBs in the bodies of the two fish is <u>high</u> enough to affect their ability to reproduce, the amount is <u>not</u> enough to affect the adult fish in any <u>other</u> way.

At this point in the scope instrument, the respondent was reminded that the fish were not

in danger of becoming extinct and that many substitutes of these species were available elsewhere

along the California Coast.

> As I mentioned earlier, these fish are not in danger of becoming extinct
> because of the millions of White Croaker and Kelp Bass along the California
> Coast that are not having reproduction problems.

A parallel statement about the fish was not included in the base instrument at this point because

respondents had already been reminded that the fish were not in danger of extinction in text

unique to that instrument.

The next part of the narrative (in both base and scope) provided assurances that the fish

injuries do not threaten human health. Early pretesting had identified this as an important concern

that needed to be explicitly addressed in the scenario and we had worked on how to do this

throughout the instrument development process. This is the first of several places where an

assurance was explicitly provided.

> <u>Some</u> people are concerned that eating White Croaker or Kelp Bass
> contaminated by these chemicals might harm <u>humans</u>. This is an important
> question, so the scientists have studied it carefully.

> SHOW MAP 3 AGAIN

> They have found that the amount of DDT and PCBs in these two types of fish is so
> small that people would have to eat fish from this <u>one</u> area (⬆ TO AREA
> MARKED IN RED) on a regular basis to be harmed. Fortunately, commercial
> fishing companies do <u>not</u> catch Kelp Bass, and the State has <u>banned</u> all
> commercial fishing for White Croaker in that area. Thus, the affected fish are <u>not</u>

See A-8 Appendix A.1, p. 8.

See the discussion of questions A-16 and W-7 in this chapter.

sold in markets or restaurants.

The State has <u>also</u> issued notices to local fishermen warning them about eating White Croaker and Kelp Bass caught there, and this warning is posted on signs. <u>Therefore</u>, it is extremely unlikely that these fish could cause <u>any</u> harm to humans. (STOP)

The next part of the narrative enhanced the plausibility of natural recovery by referring to other species affected by the deposit that had already recovered. It then explained how the natural process has worked. This explanation was illustrated by Map 4 which depicted how new uncontaminated sediment had begun to cover the contaminated layer beginning in the mid-1970's. This description was identical for base and scope with the exception of the speed with which the new sediment was described as covering the contaminated sediment. In base, the contaminated sediment was described as being buried by one foot of uncontaminated sediment by 1994; in scope, the 1994 depth was given as two feet. The greater depth in the scope version was required to describe the faster (15 years versus 50 years) natural recovery period.

Fifteen <u>years</u> ago, the deposit of DDT and PCBs was <u>also</u> causing reproduction problems in several <u>other</u> species that <u>sometimes</u> feed in the area. However, these <u>other</u> species gradually recovered and now reproduce normally.

Their recovery over the past 15 years was the result of a <u>natural</u> process. This process gradually covers the contaminated sediment on the ocean bottom with new sediment that is <u>uncontaminated</u> by DDT and PCBs. The deeper the contaminated sediment is buried, the more these chemicals are removed from the food these species eat.

```
SHOW MAP 4
```

This drawing shows how this natural process works. These little dots ⬆ are things like sand and dirt that fall through the water and settle on the bottom. The orange layer is the sediment contaminated with the DDT and PCBs ⬆.

```

```

🛑 UNTIL R IS FINISHED LOOKING AT MAP 4

Once the flow of DDT and PCBs into the sewers was stopped in the 1970s ⬆, a layer of new, uncontaminated sediment began to cover the contaminated layer. By 1994 ⬆, it had buried the contaminated layer about one foot [*two feet*] deep ⬆. This is deep enough so that there are no DDT and PCBs left in the water.

Although these chemicals now no longer affect other species, they continue to affect the four species I told you about. These species are more exposed to these chemicals than the other wildlife because of their feeding habits. [*This is because, unlike the other species, everything they eat comes from the ocean bottom in this one place where they live year round.*] Once the chemicals are buried deeper under clean sediment, these four [*two fish*] species will also recover.

The program to accelerate recovery is presented first, followed by the natural recovery option.

Until recently, there was no way to speed up this natural process. However, a procedure has now been developed to cover chemical deposits like this. A proposal has been made to use this procedure here, to speed up the recovery of the four fish and bird [*two fish*] species I told you about. The State wants to find out how people feel about this. 🛑

The next map illustrates how new, clean sediment would be placed on the ocean floor without disturbing the existing sediment, a concern expressed by some respondents in our pretests.

SHOW MAP 5

This picture shows how a speed-up program would work. This is the existing layer of sediment that covers the contaminated layer one foot [*two feet*] deep. A boat like this ⬆ would drop three [*two*] more feet of new, clean sediment down to the ocean floor without disturbing the sediment already there. This would cover the contaminated sediment under a total of four feet of clean sediment ⬆. Once they are covered by four feet of clean sediment, the DDT and PCBs would be removed from the food these species [*the White Croaker and Kelp

Bass] eat. This is because none of the animals they eat live this far beneath the ocean floor.

To enhance the credibility of the program to cover the contaminated sediments and to avoid having respondents believe the accelerated recovery program might benefit the local economy by creating jobs, the Army Corps of Engineers is described as the agency that would carry it out. The total time period until recovery to baseline conditions is described as five years.

> The <u>State</u> would pay the cost to drop the three [*two*] feet of clean sediment on the contaminated location. This program would be carried out by the <u>Army Corps of Engineers</u> which has successfully done this elsewhere. It would take <u>one</u> year to complete. Once this is done, it will take <u>four</u> more years for the animals I told you about to reproduce normally. So, within <u>five</u> years, these fish and birds [*these fish*] would be reproducing <u>normally</u>.

The narrative was broken at this point by question A-13 which gave the respondent the opportunity to ask questions about how the accelerated recovery program would work. Those who said they had questions were asked a follow-up, open-ended question, A-13A.

A-13. Do you have any questions about how this would work?

A-13A. What are they? [OPEN-ENDED]

A box in the interview provided the interviewer with clarifying answers to two questions which some respondents had asked during our pretesting: why doesn't the State remove the sediment from the ocean floor instead of covering it? and, where does the sediment come from? Here, as elsewhere, if the interviewers were asked questions for which they had not been provided with answers, they were instructed to tell the respondents: "I don't know the answer to that

---

This, and questions A-7 and A-15, served as a check on understanding and acceptance (Arrow, p. 4609) of the scenario itself.

See page 16 (base) and page 13 (scope) in Appendices A.1 and A.2, respectively.

question, but I will write it down because the researchers want to know what questions people

have about this."

The interviewer then described the natural recovery process.

A-14. If the State does not implement this program, nature will do the same
thing, but it will take longer, 50 [15] years instead of 5. This drawing shows how
this will happen.

SHOW MAP 6

Map 6 was a diagram that showed how the contaminated layer would be gradually

covered with four feet of uncontaminated sediment over a fifty year period.   In the scope

instrument, the diagram illustrated the same process but over a fifteen year period.   The

interviewer pointed to various parts of the diagram during the narrative.

This is 1994 ⬆. Over the coming years, as the new, uncontaminated sediment
continues to fall, the contaminated layer will get buried deeper and deeper.
⬆   Fifty [Fifteen] years from now, around the year 2044 [2009], the
contaminated sediment will be buried under four feet of clean sediment ⬆. As
I mentioned, this far under the ocean floor, the DDT and PCBs would be
removed from the food the four fish and bird [two fish] species eat.

The two fish and two bird [two fish] species I told you about will then have fully
recovered from their reproduction problems. Thus, instead of the 5 years it
would take for these species to recover if the State implements the speed-up
program, with natural processes it would take 50 [15] years ⬆. That is, an
additional 45 [10] years.

The next question again solicited respondent questions with a follow-up, open-ended

question.

A-15. Is there anything else you would like to know about either the speed-up

Q and A's Not in Questionnaire card. See Appendix A.3.

program <u>or</u> the natural recovery process?

A-15A.   What else would you like to know? [OPEN-ENDED]

To avoid having the respondents think this particular program must have value because the State had chosen to interview them about it, the respondent is reminded that he or she is one of many people who are being asked about <u>various</u> types of programs. Then the respondent was told that we want to know how he/she would vote if the program were on the ballot in a California referendum.

A-16. I mentioned earlier that the State has asked people about <u>various</u> types of new programs. We are <u>now</u> interviewing people to find out how they would vote if <u>this</u> program to speed up recovery were on the ballot in a California election.

The payment vehicle for this study was the California income tax. The payment frequency was a one-time payment that would be in addition to what the respondent already paid in state income taxes. Our use of a one-time household payment emphasized the monetary obligation of the respondent and is conservative relative to any payment plan that would allow the household to pay over the course of several years. The assurances that this would be the only payment and that it would go into a special fund helped to address respondent concerns, revealed in our pretesting, that the State would continue the payment indefinitely and/or use the money for other purposes.

Here's how it would be <u>paid</u> for. California taxpayers would pay a <u>one time</u> <u>additional</u> amount on their <u>next</u> year's state income tax to cover the cost. This is the <u>only</u> payment that would be required. It would go into a special fund that could <u>only</u> be used for the program to cover the contaminated sediment. The program would only be carried out <u>if</u> people are willing to pay this one time

As noted in Chapters 4 and 5, the referendum format is the elicitation framework recommended by the NOAA Panel (Arrow, p. 4608).

If a respondent who has had taxes withheld from a paycheck asks whether this additional tax would be withheld from the paycheck, the interviewers were instructed to say "yes" (Westat, 1994a; p. 4.57).

additional tax. (STOP)

At this point in the narrative, immediately before the willingness-to-pay questions, the interviewer summarized the object of choice. Possible reasons to vote against the program were also presented to enhance the credibility of the choice and to reinforce previous assurances that a vote against the program is an acceptable answer.

> There are reasons why you might vote for the speed-up program and reasons why you might vote against.

> The speed-up program would make it possible for each of the four species [*two fish species*] to reproduce normally in the South Coast 45 years [*in the place near Los Angeles 10 years*] earlier than if natural processes take their course.

The reasons to vote against were listed on a card to enhance their communication and emphasize their importance. The first reason reiterates that the injuries are restricted to the four [*two*] species of wildlife in the South Coast and that they are reversible within a 50 [*15*] year time span. The reiteration of reversibility, a major theme in the scenario, is a forceful reminder of substitute commodities as recommended by the NOAA Panel (Arrow, pp. 4-57).

> On the other hand,

> ┌─────────────────────────┐
> │ SHOW CARD H [*E*]   │
> └─────────────────────────┘

> this deposit does not harm humans and the four [*two fish*] species will recover anyway in 50 [*15*] years. (STOP)

In the question-by-question instructions for this part of the interview, the interviewer's manual reminded the interviewers about the importance of presenting this and the following material — which includes the willingness-to-pay questions — in a neutral tone and giving the respondent as much time as he/she wants to examine the material and answer the questions (Westat, 1994a; p. 4.57).

The second reason to vote against explicitly reminds respondents that there may be other issues that are more important to them that may compete with any money they might want to spend on the accelerated recovery program.

Your household might prefer to spend the money to solve <u>other</u> social or environmental problems <u>instead</u>. (STOP)

The third reason is that the amount may be more than the household wants to spend for what the accelerated recovery program would accomplish. This wording was chosen to make the respondent feel comfortable choosing to vote either *for* or *against* the program even if the respondent believes others regard it as socially desirable to vote a particular way.

Or, the program costs <u>more</u> money than your household wants to spend for this. (STOP)

REMOVE CARD H [*E*]

§ 6.4 Section W — Choice Questions

The next section begins by telling respondents how much the program would cost their household. Respondents were randomly assigned one of five versions of the questionnaire which differed <u>only</u> by the tax amount (*i.e.*, $10, $25, $80, $140, or $215) the household would pay if the program were to be approved. This dichotomous choice (*for* or *against*) for a particular level of taxation is recommended by the NOAA Panel (Arrow, p. 4612).

As noted earlier, a number of alternative public goods were specifically brought to the respondents' attention at the beginning of the survey (see questions A-1 and A-2 in section 6.2).

The same tax amounts and split-sample methodology were used in the base and scope versions.

At present, the program to speed up the covering of the contaminated sediment is estimated to cost your household a total of $(ONE OF FIVE AMOUNTS). Your household would pay this as a special one time tax added to next year's California income tax.

The interviewers were told that "household" has the same meaning as it had in the Household Screener and that if the household had more than one person who paid California income tax, the amount would be split among the taxpayers in the household. If the respondent asked a question about this, the interviewers were instructed to say: "Think of this amount as a total amount for your household" (Westat, 1994a; p. 4.59).

We developed a skip record, which folded out from a back page of the questionnaire, to help the interviewers accurately recall the respondent's vote pattern for use in later places in the interview where either the question sequence depended upon how the respondent had voted or the question required the interviewer to insert the highest amount the respondent had previously voted for.

The first choice question, W-1, asked the respondent to make a decision about the object of choice — to vote *for* or *against* the accelerated recovery program given the specified cost to his/her household. To make the decision as realistic and as immediate as possible, the choice was posed in terms of an election being held "today".

```
UNFOLD SKIP RECORD
```

See Appendix C.2.1.

The interviewer's manual (Westat, 1994a; p. 4.57) warned the interviewers that a few respondents may look to them for cues as to how they should vote at this point, and reminded them that "in fact, it doesn't matter at all whether people say `yes' rather than `no' to these questions or vice versa." The interviewers were instructed to use a neutral tone and an unhurried manner.

W-1. If an election were being held today and the total cost to your household would be a one time additional tax of $(ONE OF FIVE AMOUNTS), would you vote for the program to speed up recovery or would you vote against it?

For the reasons described in Chapter 4, two answer categories were explicitly offered to the respondent: *for* and *against*. In order to avoid the possibility of pressuring respondents who don't have an opinion at this point, the interviewers were trained to accept other responses, such as "don't know," "not sure," or "would not vote," as valid answers for this question and to record them as "not sure" without probing (Westat, 1994a; p. 4.59).

The interviewers were also told to handle any attempts by the respondent to ask them what they (the interviewer) thought about the question (W-1) by saying:

We want to know what you think. Take as much time as you want to answer this question. (PAUSE) We find that some people say they would vote for, some against; which way would you vote if the plan cost your household $____? [Westat, 1994a; p. 4.59]

Depending on how the respondent said he or she would vote at W-1, the interviewer asked a follow-up, choice question about a higher (W-2) or lower (W-3) amount. The amount was lower for respondents who said *against* or *not sure* to W-1; and, higher for respondents who said *for*. As shown in Table 6.1, each of the five tax amount versions used a different set of follow-up amounts.

Table 6.1 Tax Amounts by Version and Question

Version	W-1	W-2 (if yes to W-1)	W-3 (if no/not sure to W-1)
1	$10	$25	$5
2	$25	$45	$10
3	$80	$140	$45

| 4 | $140 | $215 | $80 |
| 5 | $215 | $360 | $140 |

The introduction to each of the follow-up WTP questions explained that it was being asked because "engineering cost estimates" could be different than originally thought, a rationale respondents generally found plausible in our pretesting.

> W-2. It is possible that the final engineering cost estimates for the program would be higher than this. If this turns out to be the case and your household would have to pay a one time additional tax of $(ONE OF FIVE HIGHER AMOUNTS) instead of $(AMOUNT GIVEN AT W-1), would you vote for or against the program?

> W-3. It is possible that the final engineering cost estimates for the program would be lower than this. If this turns out to be the case and your household would have to pay a one time additional tax of $(ONE OF FIVE LOWER AMOUNTS) instead of $(AMOUNT GIVEN AT W-1), would you vote for or against the program?

Depending on the respondents' answer to the W-1—W-3 question sequence, they were asked an appropriate follow-up question to ascertain why he/she voted that way, a procedure specifically recommended by the NOAA Panel (Arrow, p. 4609). Those who voted *against* for both the first and second, follow-up WTP question were asked:

> W-4. Did you vote against the program because it isn'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?

We chose this way of asking why the respondent voted against to encourage respondents to feel comfortable giving answers they might otherwise be hesitant to offer such as that they couldn't afford to pay for the program. The interviewers had a specific instruction at this point to record

---

The interviewers were told that the follow-up choice question "will enable researchers to get a more accurate picture of the amount of money, if any, a person would be willing to spend for the recovery program" (Westat, 1994a; p. 4.59).

verbatim all "other" answers to W-4.

All respondents who answered questions W-1—W-3 with "don't know" or "not sure" were asked W-5:

W-5. Could you tell me why you aren't sure? (BE SURE TO PROBE)

If the respondent's answer to this open-ended question was vague, the interviewer was instructed to use a probe such as: "Can you tell me what it is about the program that made you unsure?"

Every respondent who said he or she would vote *for* the program at either W-1 or W-3 was asked W-6. This question was worded to assess as specifically as possible, without leading the respondent to give one answer or another, why the respondent's household would be willing to pay the proposed amount.

W-6. People have different reasons for voting for the program. Can you tell me what covering the contaminated sediments would do that made you willing to pay for it?   [OPEN-ENDED]

Our pretesting revealed that even when the question was worded in this way, respondents sometimes gave answers that left out information they assumed was obvious to the interviewer from the context of the interview. For example, when some respondents answered W-6 by saying "help the wildlife", further probing showed the wildlife they had in mind were the species described in the interview as affected by the deposit. In order to clarify such vague answers as "help the wildlife", the interviewers were trained to use neutral and nondirective probes whenever

---

The trainers instructed the interviewers on this point as follows: "Remember, we want to hear what the respondent has to say, so keep your probes **nondirective** so you don't lead the respondent and **neutral** so you avoid biasing the respondent's answer in some direction. As always, there are no right or wrong answers. We just want to know what the respondent had in mind about this. **It is quite possible the respondent doesn't have anything in mind about this.** If so, the verbatim, after appropriate probing, will make this clear" (Westat, 1994b, p. 8.8).

respondents gave answers that were not responsive to the question (what would the program do?) or were vague or unspecific. The acceptable probes, the wording for which was provided in the questionnaire for the interviewers' ease of use, were: "Can you be more specific about what you have in mind?", "Anything else?", and "What would (covering the contamination/fixing the problem) do that made you willing to pay for the program?"

After respondents had reflected on the reasons why they voted as they did, they were offered the chance in W-7 to change their vote from *for* to *against*. This, the first of two reconsideration opportunities, was asked of everyone who voted *for* the program at either W-1 or W-3. As previously mentioned, we paid special attention in the scenario to neutralizing a concern held by some respondents that the deposit could harm human health. W-7 specifically raised this concern to make it clear to respondents who continued to harbor it that the program would only speed up the recovery of the affected wildlife. Respondents who had second thoughts and wanted to change their *for* vote for any other reason could also take advantage of this reconsideration opportunity.

> W-7. It is not unusual for some people to vote for the program because they are concerned that these DDT and PCBs may harm human health. Suppose human health is definitely not affected in this situation and the program would only speed up the recovery of these four species of fish and birds [*two species of fish*]. Would you vote for or against the program if it cost your household $[LARGEST AMOUNT RESPONDENT VOTED FOR]?

§ 6.5 Section B — Perception of Injury, Program, and Interview

The first set of questions in Section B asked the respondents what they had in mind or had assumed about some of the scenario features when they voted on the program. For example, did

---

The answers to these questions are another type of check on respondent understanding and

they believe that natural processes would take about fifty (or, fifteen in scope) years to return things to normal as they had been told? Questions requiring this type of introspective assessment may be unfamiliar to respondents. Answers to such questions, nevertheless, can help us check which features were accepted by the respondents when they voted.

The introductory statement and the first question, B-1, were worded to convey the request for this type of information as clearly and respectfully as possible.

> I gave you a <u>lot</u> of information <u>before</u> you voted. Please think back to a few moments ago <u>when</u> you decided how to vote. We are interested in what you were thinking then.
>
> B-1. <u>First</u>, did it seem to you that DDT and PCBs <u>could</u> cause the reproduction problems I told you about?

The next question in this sequence asked whether the respondent had *accepted* the fifty (fifteen) year natural recovery period as plausible; for this reason, the wording we used was "seem to you". The follow-up question, B-3, asked those who said "no" if they thought it would take a "lot more" or a "lot less" than 50 (15) years. The interviewers were instructed to record verbatim any other type of answers the respondent gave to B-3.

---

acceptance of the scenario (Arrow, p. 4609). Each of the several methods we use to check understanding and acceptance has drawbacks; taken together they provide useful information.

Based on our pretest experience with this survey, the interviewers were told that "sometimes respondents wonder why they are being asked questions like this about what they were thinking when they answered the vote questions — after all, they might say, you told them that it would take fifty (fifteen) years, why should they doubt it." The interviewers were instructed to tell such respondents: "We find that some people have different ideas about this. It is important for us to know what <u>you</u> had in mind." (Westat, 1994a; p. 4.71)

In earlier presentations of this material in our pretesting, some respondents resented these questions because they took them to be a quiz.

Our development work suggested that words like "plausible" and "reasonable" were not widely enough understood in this context to use in wording these questions.

B-2. When you decided how to vote, did it seem to you that natural processes would take about fifty [15] years to return things to normal?

B-3. Did it seem to you that it would take a lot more than 50 [15] years or a lot less than 50 [15 years]?

Questions B-4, B-5, and B-6 explored the respondent's assumptions about the

effectiveness of the program and the payment period.

B-4. When you decided how to vote, did it seem to you that the speed-up program would be completely effective in solving the reproduction problems within five years?

SHOW CARD I [F]

B-5. Did it seem that the program would be . . . mostly effective, somewhat effective, not too effective, or not effective at all?

B-6. When you decided how to vote, did you think your household would have to pay the special tax for the program for one year or for more than one year?

The next question asked whether the respondent felt pushed to vote one way or the other

by the interview. For those who felt they had been pushed one way or the other, two follow-up

questions (B-7A, B-7B) probed which direction and why they felt this way.

B-7. Thinking about everything I have told you during this interview, overall did it try to push you to vote one way or another, or did it let you make up your own mind about which way to vote?
B-7A. Which way did you think it pushed you?

B-7B. What was it that made you think that? (PROBE: "Can you be more specific about what you have in mind?" "Anything else?")

Question B-8 asked for the respondent's assessment of the seriousness of the injuries

---

Card I contained the answer categories for B-5. See Appendix A.1.

Question text that follows ". . ." was presented in the questionnaire as lower case answer categories (interviewers were instructed not to read anything that appeared in upper case); a NOT SURE answer category was also included but not read out loud. See Appendix A.1.

described in the scenario.

┌─────────────────────────────┐
│ SHOW CARD J [G]   │
└─────────────────────────────┘

B-8. All things considered, would you say the fish and bird [*fish*] reproduction problems I told you about in the South Coast are . . . not serious at all, not too serious, somewhat serious, very serious, or extremely serious?

§ 6.6 Section B — Household Recreational Activities

The next eight questions asked about various types of household recreational activities. When five answer categories were used, show cards were used to display the categories; see Appendix A.1 for show cards and the answer categories for questions asked without the aid of a show card.

Now I would like to ask you a few questions about your household's recreational activities.

B-9. In the past five years has anyone in your household gone fishing?

B-10. Is that saltwater fishing, freshwater fishing, or both?

B-11. In the past five years has anyone in your household gone boating?

B-12. Is that saltwater boating, freshwater boating, or both?

B-13. Does anyone in your household like to identify different species of birds?

Card J listed the five answer categories. See Appendix A-1.

SHOW CARD K [*H*]

B-14. How often do you personally watch television programs about animals and birds in the wild . . . very often, often, sometimes, rarely, or never?

SHOW CARD K [*H*] AGAIN

B-15. How often do people in your household go to the beach at the ocean . . . very often, often, sometimes, rarely, or never?

SHOW CARD K [*H*] AGAIN

B-16. How often do people in your household eat fish . . . very often, often, sometimes, rarely, or never?

The last question in Section B asked respondents to indicate the degree to which they thought of themselves as an "environmentalist". If they asked the interviewer what was meant by this term, they were given a standard survey reply to such questions: "Whatever it means to you." (Westat, 1994a; p. 4.83)

SHOW CARD L [*I*]

B-17. On another subject, would you say you think of yourself as an . . . environmental activist, a strong environmentalist, a somewhat strong environmentalist, a not particularly strong environmentalist, or not an environmentalist at all?

§ 6.7 Section C — Respondent Household Experience and Demographic Characteristics

The first series of questions in Section C obtained information about the respondent's household and personal characteristics.

The Household Screener which the interviewer had previously administered to select the main

Now, I have just a few questions about your background.

C-1. First, in total, how many years have you lived in California?

C-2. Have you ever been to Catalina or any of the other Channel Islands?

Those who said "yes" to C-2 were asked the next question.

C-3. Was your most recent visit within the past five years?

Everyone was asked C-4.

C-4. Do you intend to move outside California in the next few years?

The next question, C-5, was only asked of respondents who were interviewed in Los Angeles or

Orange Counties.

C-5. Do you intend to move outside (L.A./Orange) County in the next few years?

C-6. In what month and year were you born?

The interviewer coded the respondent's answer to the education question, C-7, into one of eleven categories ranging from "through 8th grade" to "doctorate degree". C-7. What is the highest year of school you completed or the highest degree you received?

C-8. During 1993, how many adults in your household, including yourself, worked for pay?

C-9. How many people live in this household who are younger than 18?

C-10. Do you have children of any age who live outside this household?

C-11. Do you have any grandchildren?

Because pretests showed that some respondents did not include retirement income when

they answered the income question, we asked C-12 as a way to remind them of this type of

interview respondent also collected additional information. See Appendix C.2.1.

income.

>C-12. Did anyone in your household have any income from social security or pensions in 1993?

In the next question, the interviewer had the respondent report his or her household income from categories listed on a card, a standard survey research device.

>C-13. I'd like you to think about the income received last year by everyone in your household.

>SHOW CARD M [J]

>Adding together all income for everyone in your household, which letter on this card best describes your household's total income for last year — 1993 — before taxes?   Please include wages or salaries, social security or other retirement income, child support, public assistance, business income, and all other income.

The next question was asked to identify respondent households that did not have to pay any California income tax.   Because the income range covered by the two lowest income categories — $0 to $19,999 — included virtually all California households who might not owe California income tax, C-14 was only asked of households in these two income categories (A or B).

>C-14. Did anyone in your household pay any California income taxes for last year, 1993, either by having taxes withheld from wages, retirement income, or other money received, or by sending money to the State with a tax form?

All respondents were asked about their future household income prospects.

>C-15. If things go as you expect, do you think your household's total income for this year will be about the same as last year, higher than last year, or lower than last year?

This card listed 11 income categories ranging from "under $10,000" to "$100,000 or more." See Appendix A.1.

§ 6.8 Section C — Strength and Reassessment Questions

This series of questions was <u>only</u> asked of respondents who had said they would vote <u>for</u> the program at one of the amounts in the W-1 to W-3 question sequence and those who had <u>not</u> changed their vote at W-7. They were asked how difficult it would be for them to pay the highest amount they voted *for* and how strongly they favored the program at this cost. Everyone who reported either: (1) that it would be "very difficult" or "somewhat difficult" for his/her household to pay the amount, or (2) that he/she favored the program "not too strongly" or "not at all strongly" was offered a chance to reconsider his/her vote. As displayed below, some questions were repeated in the questionnaire. This repetition was to make the skip patterns more manageable; no respondent was actually asked the same question more than once.

C-16. Now that we're close to the end of the interview and you have been able to think a bit more about the situation, I'd like to give you a chance to review your answers to the voting questions.

You said you would vote for the program to speed up the recovery of the four fish and bird species [*the two fish*] if it cost your household a one time additional tax payment of $_____.

C-17. How <u>difficult</u> would it be for your household to actually pay $[LARGEST AMOUNT RESPONDENT VOTED FOR] next year if the program passed? Would it be . . . very difficult, somewhat difficult, not too difficult, or not difficult at all?

C-18. Now that you have had a chance to think a bit more about this, would you vote "For" or "Against" the program if it cost your household $[LARGEST AMOUNT RESPONDENT VOTED FOR]?

C-19. How <u>strongly</u> do you favor the program if it cost your household this much money? Would you say . . . very strongly, strongly, not too strongly, or not at all strongly?

C-20. Now that you have had a chance to think a bit more about this, would you vote "For" or "Against" the program if it cost your household $[LARGEST AMOUNT RESPONDENT VOTED FOR]?

C-21. How <u>strongly</u> do you favor the program if it cost your household this much money? Would you say . . . very strongly, strongly, not too strongly, or not at all strongly?

C-22. Why is that? [OPEN-ENDED]

§ 6.9 Section C — Miscellaneous Questions

The following question was only asked in the scope version. It served to inform scope respondents that the particular set of injuries described to them in the scenario was not definitive.

[scope only] C-23. In this interview I described the effects of DDT and PCBs on the White Croaker and Kelp Bass that live off the Los Angeles coast. Some scientists think DDT and PCBs may still be causing reproduction problems in two other species in the South Coast. These are the Bald Eagle and the Peregrine Falcon. If this turned out to be the case, would you consider the problem caused by these chemicals to be more serious?

In order to measure respondent attitudes about different institutions and groups, we asked the following questions.

[both base and scope] C-24. I'd like to know how much confidence you have in some of the institutions and groups in this country.

SHOW CARD N [K]

First, (READ X'd ITEM) . . ., generally speaking, would you say you have a great deal of confidence, some confidence, hardly any confidence, or no confidence at all in . . .? (READ EACH ITEM, BEGINNING WITH X'd ITEM; CIRCLE ONE CODE FOR EACH; RE-READ STEM AS NECESSARY.)

    a. University scientists
    b. U.S. Congress
    c. Scientists who work for industry
    d. Newspapers
    e. California state government
    f. Large corporations

This card listed the four answer categories.

Because the State of California was identified as the sponsor of the survey, we asked an additional question about how much trust the respondent had in the state government.

SHOW CARD O [*L*]

C-25. How <u>much</u> of the time do you think we can trust the California state government to do what is <u>right</u>? Would you say . . . always, almost always, most of the time, some of the time, almost never, or never?

The following question measured which method of paying for environmental improvements, higher prices or higher taxes, the respondent prefers. In our development work we learned that respondents volunteered two other answers with some frequency, so we included those among the answer categories that were not read to the respondent. They were: "neither" and "don't care which one".

C-26. There are different ways for people to pay for new programs to protect the environment. (STOP) One way is for the government to pay the cost. This will raise everyone's taxes. (STOP) The other way is for businesses to pay the cost. This will make prices go up for everyone. If you had to choose, would you prefer to pay for new environmental programs . . . through higher taxes, or through higher prices?

The next questions in this survey concerned what languages the respondent spoke at home and whether he or she was a United States citizen. Only respondents who answered "no" to C-27 were asked C-27A.

C-27. Do you usually speak English at home?

C-27A. What language do you usually speak at home?

C-28. Are you a citizen of the United States?

The last few questions are standard items that Westat asks in surveys of this type for validation purposes and to record respondent characteristics that can be observed by the interviewer.

> C-29. What is your full name and phone number, in case my supervisor wants to check my work? (RECORD FULL NAME AND PHONE NUMBER ON RECORD OF ACTIONS. DO NOT RECORD IT HERE.)
>
> INTERVIEWER, PLEASE NOTE THE FOLLOWING ABOUT THE RESPONDENT BY CIRCLING THE NUMBER OF THE CORRECT RESPONSE:
>
> C-30.   SEX
>
> C-31.   RACE
>
> C-32.   ENTER R'S ZIP CODE:
>
> C-33.   ENTER R'S PSU #:

§ 6.10 Section D — Interviewer Evaluation Questions

The interviewers were asked to give their impressions about certain aspects of the interview by filling out the questions in Section D. We were particularly interested in any information they might provide about any difficulty the respondent might have had in understanding the material. All questions in this section were answered by the interviewers after they left the respondents' homes. The interviewers were told:

> Section D of the questionnaire is designed to provide us with feedback from all interviews. It is crucial to the evaluation effort that you answer all applicable questions as fully as possible. You, as an interviewer, are our most important source of information for evaluating these topics. [Westat, 1994a; p. 4-107]

The race categories, based on the 1990 census categories, were as follows: White, Not Hispanic; White, Hispanic; Black, Not Hispanic; Black, Hispanic; Asian; and Other.

Questions D-1 to D-6a asked for the interviewers' impression about the interview situation, how

the respondent attended to the interview, and the difficulties the respondent may have had.

D-1. What was the reaction of the respondent as you read through the material beginning with A-3 through A-16? (This is the descriptive material including the maps and charts.)

   a. How distracted was the respondent?
   b. How attentive was the respondent?
   c. How well did the respondent understand this material?

D-2. Did the respondent say anything suggesting that he or she had any difficulty understanding either the natural recovery process or the speed-up program?

D-2A. Describe the difficulties. [OPEN-ENDED]

D-3. Did the respondent have any difficulty understanding the vote questions (W-1 through W-3)?

D-3A. Describe the difficulties. [OPEN-ENDED]

D-4. When you asked the voting questions did you feel the respondent was impatient to finish the interview?

D-4A. How impatient was the respondent?

D-5. How serious was the consideration the respondent gave to the decision about how to vote?

D-6. Not counting you and the respondent, was anyone age 13 or older present when the respondent voted?

D-6A. Do you think the other person(s) affected how the respondent voted or don't you know?

The final question invited the interviewers to make any other comments they wished about

---

The scale included the following categories: extremely, very, somewhat, slightly, not at all, and not sure.

The answer categories were very impatient, somewhat impatient, a little impatient, not very impatient, and not sure.

The answer categories were extremely serious, very serious, somewhat serious, slightly serious, not at all serious, and not sure.

the interview. No specific instruction was given about this except that they should "record here any other comments you think would be useful about how the interview worked and how the respondent `took' the interview." (Westat, 1994a; p. 4.111]

D-7. Do you have any other comments about this interview? [OPEN-ENDED]

§ 7 Main Survey Administration

§ 7.1 Introduction

Westat's administration of the main survey consisted of several distinct steps. A random sample of California blocks was drawn, the individual dwelling units in those blocks were listed, and a random sample of the listed dwelling units was selected. An interviewer's training manual was prepared, and Westat's interviewers attended a three-day training session to ensure consistent and proficient administration of both the base and scope versions of the survey instrument. During the five months of main survey data collection, the interviewers were supervised by regional field supervisors and a field manager. As the interviews were completed, Westat conducted quality control edits and validations. At the end of the data collection, sample weights were constructed. Finally, data sets containing the responses to both close-ended and open-ended questions were prepared. This chapter details each of these steps. As will be seen, three of the NOAA Panel's recommendations implemented in our study pertain directly to survey execution: in-person interviews, a probability sample, and, to the extent possible, minimization of non-response.

Westat, headquartered in Rockville, Maryland, is one of the country's most respected survey research firms. A copy of Westat's corporate brochure can be found in Appendix C.1.

See Chapter 5 for a discussion of the base and scope instruments and Appendices A.1 and A.2 for copies of the base and scope survey instruments, respectively.

§ 7.2 Sample Design

Westat designed the main study sample 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). A multi-stage area probability sample was designed to give each residential dwelling unit in California an equal chance of selection. The selection of the sample followed standard procedures for multi-stage area frame designs that have been used for decades by high-quality survey organizations.

At the first stage of selection, all the counties in California were assigned to Primary Sampling Units (PSU's). Many of the PSU's consisted of multiple counties, some of single large counties, and Los Angeles county was divided into two PSU's (the city and the rest of the county). Thirteen PSU's were then selected with probabilities proportional to their 1990 Census population counts. Within the selected PSU's, 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.

The Census Bureau's definition of a dwelling unit (DU) was used: a house, an apartment, or group of rooms or a single room occupied as separate living quarters (that is, the occupants do not live and eat with any other person in the structure, and there is direct access from the outside or through a common hall or area). See Westat (1994a).

See Kish (1965).

These were as follows: 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.

§ 7.3 Selection of Dwelling Units

From August 18 to October 9, 1993, Westat's trained listers canvassed the 652 selected segments and listed every dwelling unit (DU) they found. (For those segments with a very large number of DU's, only a "chunk" chosen by Westat's sampling department, with probabilities proportional to its size, was listed.) A random selection of dwelling units was then drawn from the listed DU's by Westat statisticians. The number selected (4,800) was determined after estimating rates of occupancy (some DU's will be vacant), eligibility (some won't contain English-speaking adults), and response (some won't cooperate with the request for an interview) so as to yield approximately 3,000 interviews.

At the start of the main study data collection in March, interviewers followed a prescribed probability procedure to sample DU's not recorded by the listers. This procedure corrected, in an unbiased manner, for DU's missed by the listers as well as for any units constructed after the listing was conducted. It produced 21 additional DU's. Thus the total sample consisted of 4,821 dwelling units. These 4,821 DU's were randomly assigned within segments (in a 2 to 1 ratio) to the base and scope samples.

This procedure is described in Westat's Listing Manual for this study.

The listing process revealed that one of the selected segments contained no dwelling units; hence, the selected DU's come from 651 segments.

See Appendix C.2.1.

Cases were also randomly assigned to the five tax amount versions described in Chapter 6.

§ 7.4 Interviewer Training

The 59 professional interviewers participating in the study attended a three-day training session held on March 5-7, 1994, in San Diego, CA. All of the interviewers had prior household interviewing experience. The training session was conducted by the study's Project Manager, Susan Rieger, assisted by the Field Director, Field Manager, and three Regional Field Supervisors. The interviewers, field supervisors, and field manager were not informed of the survey's intended use in litigation. The study was referred to simply as the California Issues Study (CIS). Only a small number of senior staff in the Rockville office knew of the intended use and they exercised care to ensure that its purpose was not communicated to the field staff.   The interviewers had been given an initial set of study materials to read before attending training. The training consisted of scripted lectures, exercises, interactive small group sessions, and role-playing sessions (using prepared scripts) in which one trainee took the role of the interviewer and another played the role of the respondent.

After general introductions, the first morning of the training began with an overview of the CIS study. The discussion then turned to the interviewer's role and a brief description of the interviewer's materials. After a break, the training reconvened for a demonstration interview to show the way the interview was to be administered. That was followed by a detailed discussion of the first of the interviewer's tasks: locating the dwelling unit and selecting a respondent for the main interview. The remainder of the afternoon was devoted to the administration of the main interview. This took place in small groups led by the regional supervisors and the project manager. The key features of the main interview were highlighted; special emphasis was placed on using the visual aids, reading the narrative sections, and following the skip patterns.

Westat's training procedures are further described in Appendix C.5.1.

The second day of training began with a detailed comparison of the base and scope instruments. They were identified as "Version A" and "Version B" or "blue" and "yellow" (corresponding to their colors), respectively; the words "base" and "scope" were never used with any of the field staff. The interviewers were told that the two versions were being fielded because of scientific uncertainty about the number of species affected. The trainees then broke into smaller groups for two interactive sessions led by the regional supervisors. The first session included lecture and practice with probing techniques. The second session included a detailed group discussion of the screening procedures, the record of actions, and the non-interview report (NIR) form. The discussion was followed by a role-playing exercise. After lunch, the role-playing exercise was completed and the rest of the day devoted to further role-playing.

The third day began with a review of the probing exercise (a self-administered test on probing) and administrative procedures. The remainder of the day was spent in role-playing with both versions of the survey instrument. As part of concluding comments, interviewers were instructed to practice administering the two versions of the survey instrument at home before they conducted interviews at sampled DU's.

§ 7.5 Interviewer Supervision

All interviewers reported to one of the three regional supervisors, who in turn reported to the field manager. Supervisors were responsible for conferring with interviewers regularly, reporting on and managing progress, performing quality control edits, and validating interviews. Interviewers interviewers participated in conference calls with other interviewers and supervisors to share their strategies on gaining cooperation.

Supervisors entered data on interviewing production, time, and expenses into a machine-

readable file that was set up to generate field status reports. Supervisors also reported weekly by telephone to the field manager on survey progress, case assignments, and refusal conversion strategies. In addition, the field director had a weekly telephone call with the supervisors and the field manager.

§ 7.6 Quality Control Edits

Interviewers sent their completed questionnaires to their supervisors on a weekly basis. Upon receipt, supervisors were responsible for a comprehensive field edit of the questionnaires before sending them to the home office for further editing and data entry. The edits, for completeness and accuracy, used the form shown in Appendix C.3. The form covered respondent selection, skip patterns, probing, verbatim recording, and other administrative matters. Results of the edits were discussed with the interviewers.

The edits uncovered 16 cases in which respondent selection within the household was carried out improperly. None of these cases was included in the final data set; they were all counted as non-respondents to the main interview.

§ 7.7 Main Survey Data Collection

The main survey data were collected over a 24 week period, from March 8 to August 23.

To facilitate quicker turnaround of the final cases, toward the end of the field period these edits were conducted by staff in Westat's Rockville office.

In another instance, a respondent received an emergency phone call during the administration of section B of the questionnaire; the main interview was temporarily terminated. After consulting with supervisors, the interviewer returned to administer the demographic questions (C-1 through C-15) contained in section C. The section B and the section C questions not administered in the second visit were assigned a value of 9 (for "not ascertained") and the case was included in the final data set.

In the first week of the field period, the Los Angeles Times reported that small amounts of DDT were found in two backyards in an unincorporated area east of Torrance, half a mile from the former Montrose Chemical site. In subsequent weeks, the EPA discovered that the DDT was in larger quantities than expected; various government agencies held three public meetings in the Torrance area; and, more than 30 families were relocated so that the EPA could excavate the DDT-laced soil.

While none of the selected DU's were in this neighborhood, three segments (encompassing 22 selected DU's) were located in sections of Torrance just to the south of this neighborhood, and two additional segments (encompassing 13 selected DU's) were located in the city of Carson which is just to the south-east. Given the close proximity of these 35 DU's to a neighborhood where the excavation of DDT was causing intense concern about human health, no attempt was made to administer the main interview there. As a result, the population to which our results apply consists of all English-speaking California households except those near the excavation, which we defined as zipcode 90502 (the location of the excavation) and the four zipcodes sharing a common boundary with it (90501, 90248, 90710, and 90745).

In the beginning of June, an incentive program was introduced to minimize attrition of interviewers and to reward interviewers for completing the more difficult cases that remained. These cases included a large number of initial refusals and cases where it was difficult to find household members at home. The incentive plan, similar to ones Westat had used on other large studies, followed a two-tiered approach. It provided a monetary incentive for total number of cases completed during the data collection period as well as a weekly incentive for cases

---

Los Angeles Times, March 10, 1994, at B3.

completed over a set amount for the remainder of the field period.

§ 7.8 Validation of Interviews

The original plan was for supervisors to validate approximately 10 percent of each interviewer's assignment. The cases to be validated were randomly preselected in advance of the field work. Thus, both interviews and non-interviews were validated. For interviews completed after the incentive plan went into effect the validation rate was increased to 100 percent.

Validations were performed by telephone using the form shown in Appendix C.4. Validations on cases without telephone numbers were attempted in-person. Problems with interviews conducted by two interviewers were discovered. As a result, all of the cases they were assigned were validated (*i.e.*, 100 percent of the work they completed both before and after the introduction of the incentive plan). This revealed that a total of 30 interviews had not been conducted with residents of the selected dwelling unit. In another 8 cases, the validator was unable to determine whether the interview had been conducted. For many of the 30 failed validations, information about the interview topic was conveyed in the course of conducting the verification. In order to avoid self-selection bias arising from knowledge of the topic, no attempt was made to interview the correct respondent in such instances. In other cases, however, nothing about the topic was conveyed during the validation and, in two of these instances, a main interview was later conducted with the appropriate person. The remaining 36 cases were treated as "other nonresponse". (For further details, see Appendix C.5.4).

For further details, see Appendix C.5.3.

There were also 9 cases to which the two suspect interviewers had assigned various non-response outcomes
that validators could not verify. Eight of these were also treated as "other nonresponses". The

§ 7.9 Sample Completion

The outcome of the interviewers' attempts to complete a Screener—designed to collect information on household composition and to select a respondent for the Main Interview—was as follows:

---

ninth was coded a refusal, which was the outcome after another interviewer tried to conduct the interview.

A copy of the CIS Household Screener can be found in Appendix C.2.1. The other field materials (*e.g.*, refusal conversion letters, not at home/unable to contact letter, community leader letter, "Sorry I Missed You" card, "No Habla Español" card) used by the interviewers and, when appropriate, mailed to the selected addresses can be found in Appendices C.2.2 to C.2.7.

Screeners Completed.. 3,391
Not an Occupied Dwelling Unit......... 503
Language Barriers...... 152
Refusals 551

Physical/Mental Handicaps....   39   *737 eligibility*

unknown
Never Reached...........   52
Other Nonresponses...   95
Torrance Area Ineligibles   35
Other Ineligibles......... .... __3__

TOTAL.......... 4,821

The results from the 3,391 cases where a respondent was randomly selected from the Main Interview were as follows:

Main Interviews Completed... 2,810
Refusals   269
Language Barriers......   189
Physical/Mental Handicaps....   26
Never Reached..........   54
Other Nonresponses... ....__43__

  This includes 35 cases that could not be validated; 26 cases where the household moved before the Screener could be administered; 15 cases where the final outcome was unknown (*e.g.,* questionnaire was lost in the mail); 10 cases where the interviewer was unable to gain access (*e.g.,* selected DU was in a locked building); and 9 cases where the correct DU could not be identified due to insufficient listing information.

  This consists of the addresses in the five segments near the Montrose site. At the very end of the field period, attempts were made to gather Screener information about the composition of these households, but no respondent was selected to be interviewed for the Main Interview. The information from these Screeners was used in the construction of sample weights as well as to estimate the number of households in the Montrose area. This estimate was subtracted from the estimated total number of California households in order to arrive at the population to which we extrapolated our results (see Section 7.10).

  This consists of addresses occupied on a temporary basis by visitors who resided outside of California.

  This consists of 16 cases in which respondent selection within the household was carried out

TOTAL.......... 3,391

The response rate is the number of completed main interviews divided by the number of eligible households. Thus, computing the response rate involves making an assumption about the eligibility of the 737 occupied dwelling units that were nonresponses to the Screener for other than language reasons. The standard survey practice is to assume the same proportion of these cases was eligible as for those cases whose eligibility was determined (Council of American Survey Research Organizations, 1982), which in this instance is 90.3 percent. Using this approach, the response rate was 72.6 percent: 2,810 divided by [4,821 - (503 + 379 + (0.097 * 737))]. That is, in calculating the response rate, we removed from the denominator the 503 addresses that were not occupied DU's, the 379 known ineligible cases (341 language barriers, 35 Torrance area ineligibles, and 3 other screener ineligibles), and 71 additional cases representing our best estimate of the ineligibles among the screener nonresponses.

---

improperly, 15 cases in which the household moved before the Main Interview could be administered, 9 cases where the validation confirmed the Screener but not the Main Interview, and 3 cases where the Main Interview could not be administered before the end of the field period.

  Of the 3,546 occupied DU's outside of the Torrance area whose status was determined (completed screeners, screener language barriers, and screener other ineligibles), 3,202 (or 90.3 percent) were members of the eligible population (3,546 less the screener language barriers, main interview language barriers, and screener other ineligibles).

  The response rate for the base sample was 72.1 percent and for the scope, 73.8 percent, a difference that is not statistically significant. The response rates for each PSU are provided in Appendix C.6. Due to rounding, the over-all response rate is shown in the appendix table as 72.7.

  The lower-bound estimate of the response rate, assuming that all of the 737 unknown eligibility cases were in fact eligible, is 71.3 percent.

§ 7.10 Sample Weights

As information about the survey topic was not provided to individuals until after the main interview began, willingness to pay for the program to speed-up the recovery of the affected species could not have directly affected whether or not a household responded. It is possible, however, that other characteristics (*e.g.,* household size or residence in large urban areas) were related to responding/non-responding status. Thus the composition of the interviewed sample could differ from that of the total sample initially chosen. In addition, some parts of the population may not be represented in a sample either because dwelling units were missed by listers or because individuals who live in a dwelling unit were not reported as living there. This is referred to as undercoverage. Finally, the fact that samples are drawn randomly means that chance processes may cause the sample characteristics to depart from those of the population from which it was drawn. This is known as sampling variability.

In order to limit the impact of sampling variability and reduce the potential for error from nonresponse and undercoverage, sample weights were constructed following standard survey procedures. The sample weights incorporated both nonresponse adjustments and post-stratification to household totals from the 1993 Census Bureau's Current Population Survey (CPS). The nonresponse adjustments were done within groups defined by age of householder, race/ethnicity, and household type (married couple present versus other). The weights of those who responded within a group are increased by a factor that allows them to represent both themselves and the non-respondents within the group. To the extent that respondents are similar to the non-respondents within a group in terms of responses to a survey item, the potential for non-response bias in the corresponding survey estimate is reduced.

The post-stratification involved weighting the sample so it reflected the California

distribution of the 1993 CPS on age of householder, race/ethnicity, household type, and geographic area of California. The sample weights of respondents were adjusted so that aggregate totals corresponded to Census figures. This reduces variation from the chance nature of sample selection as well as adjusts for any coverage differences among the groups used for the post-stratification.

No additional corrections to the data set beyond those implied by the weighting scheme described above have been made because we have assumed that dwelling units chosen for our sample but not interviewed are missing at random with respect to their willingness-to-pay values within the groupings used for the weighting adjustments. This assumption is plausible largely because a household's decision to participate or not participate in our survey was made without knowledge of the survey's subject matter. It is possible that households who are very difficult to find at home or who generally refuse to be interviewed have systematically different willingness-to-pay values, but it is unclear whether the values might be higher or lower. In any event, our response rate is sufficiently high that any non-response effects should be reasonably small.

Due to logistical and time considerations, no foreign language versions of the final questionnaire were used. As a result, non-English speaking households were not eligible to be

---

The weights also took into account the departures from equal probabilities of selection that occurred in 6 of the 651 segments. In 4 segments, clerical errors meant that households were selected with probabilities that were too large. In 2 other segments, there had been extraordinary growth in population since the 1990 Census. As a result, these 2 segments would have contributed a disproportionately large fraction of the total sample if an equal probability design had been followed. It is standard practice in such instances to restrict the number of DU's selected, as well as to trim the weights associated with the cases so as to minimize the mean square error of the results. For further details of the weighting, see Appendix C.5.5.

As the NOAA Panel points out, response rates substantially higher than ours are unlikely to be achieved in contingent valuation surveys (Arrow *et al.*, 1993; p. 4611).

interviewed. On the basis of the characteristics of a 5 percent sample of the California households

from the 1990 Census (the Public Use Microdata Sample), Westat reduced the 1993 CPS estimate

of the number of California households to reflect the proportion of that were non-English

speaking in the subgroups used for post-stratification. This yields an estimate of 10,347,108

English-speaking California households to which our results may be extrapolated.

§ 7.11 Data Entry

As the questionnaires were received at Westat's home office, the numeric and verbatim

responses were entered into separate computer files by the data entry department. The numeric

data were entered as they appeared on the questionnaire. The data entry staff was instructed to

enter a value of "9" in those instances where the question was blank but should have been asked.

The data were entered in batches independently by two persons (that is, there was 100%

verification of the data entry). When data entry and validation activities for a batch of

questionnaires were complete, an ASCII file containing the numeric responses was electronically

mailed to NRDA. The batch of questionnaires and a diskette containing both the numeric

response data file and a verbatim response file were sent to NRDA.

If no one in the household spoke English, but someone spoke Spanish, an attempt was made to send a Spanish speaking interviewer to administer the screener to obtain household information. Once the screener was completed, if it was determined that no eligible household member spoke English well enough to be interviewed, the main interview was closed out as a language barrier. If it was not possible to send a Spanish speaking interviewer to administer the screener, the screener was closed out as a language barrier.

The total number of California households was estimated to be 11,107,204, of which 10,410,160 were English-speaking. This number was then reduced by 63,052, Westat's estimate of the number of households in the five zipcodes described in Section 7.7.

For a description of how Westat tracked each questionnaire before sending to NRDA, see Appendix C.5.2.

Questionnaires arriving at NRDA were logged and filed, and the numeric responses re-entered by NRDA staff. Once a batch was re-entered, that data set was compared with the data set provided by Westat. For each case, a direct comparison was made of the two values for each variable. Differences were reconciled by an examination of the source questionnaire; and a data set was constructed incorporating the reconciled values of the two data sets.

Using the reconciled numeric response data set, NRDA corrected skip pattern violations and recording errors. A computer program was written that assigned a value of "9" (categorized as *not ascertained* in the Appendix D.1 tables) to those questions that the respondent was not asked but should have been asked. A value of "." was assigned to those questions which the respondent was asked but should not have been asked. In addition, a separate program was written to treat the less than two percent of the interviews that contained errors at the voting questions (W-1, W-2, and W-3) and the reconsideration questions (W-7, C-17—C-20). These were either recording errors (*i.e.,* the interviewer circled the appropriate answer category on the skip record but not at W-1, W-2, W-3, or W-7) or cases where the respondent changed his or her mind about an answer and the interviewer circled a second answer category without putting a line

---

The most common differences were as follows: Westat and NRDA interpreted the handwriting differently for questions that required interviewers to record a number (*e.g.,* psu, zipcode, year of birth); Westat entered a value of "9" and NRDA followed pre-specified decision rules for questions where the interviewers circled more than one answer category (*e.g.,* respondent answered "in between 2 and 3" so interviewer would circle both 2 and 3 or respondent at W-4 would answer both 1 or 2 and make a spontaneous comment that was recorded under "OTHER [SPECIFY]" and thus coded 3); Westat's coders referred to other field material documents (whereas NRDA didn't) when the interviewer failed
to circle A.M. or P.M., failed to enter correct PSU, didn't check Box 1 or Box 7, or did not code race or sex on the main interview itself; and cases where Westat's coders neglected to enter a value of "9" for questions that were not asked but should have been.

See Appendix C.7 for copies of the recode files (executable in STATA).

through the first code that was circled. This cleaned data set was used in the analysis reported

elsewhere in this report. Tabulations of this cleaned data set, weighted and unweighted, are found

in Appendix D.1.

NRDA staff also re-entered the verbatim responses. The two verbatim response data sets

were compared by visually comparing the entries for each question. Inconsistencies were

resolved by reference to the source questionnaire, and a data set was constructed incorporating

the reconciled responses of the two compared data sets.

If the respondent changed his/her mind after answering a question, the interviewers were
instructed to put a line through the first code that was circled and write next to it "RE" (an
abbreviation for respondent error).

The discrepancies involved such things as transposed prepositions and pronouns and the linking
of spontaneous comments to a page number versus a specific question number.

§ 8 Evaluation of Qualitative Measures of Survey Reliability

§ 8.1 Introduction

This chapter examines the qualitative evidence underlying the quantitative data analysis presented in Chapter 9. In section 8.2, *verbatim* responses to open-ended questions in the base survey instrument are examined. The primary focus is on the elicitation questions recommended by the NOAA Panel that asked respondents to explain their reason(s) for voting *for* or *against* the accelerated recovery program or why they were not sure about how they would vote. While a qualification to that recommendation is noted in Chapter 4, qualitative data from the survey provide evidence that respondents paid attention to the survey and took the choice opportunity seriously, that respondents' decisions reflected their perceptions of the object of choice and their preferences for it, and that their choices were not influenced by extraneous factors, one's confidence in the reliability of the data is increased. Section 8.2 also examines the additional kinds of information respondents requested during the presentation of the injuries and the accelerated recovery program.

In section 8.3, the responses to section B debriefing questions, which provide additional information about how respondents perceived various aspects of the injuries and program, are examined. Section 8.4 explores the characteristics of those respondents who changed their *for* votes to *not for* votes when they were given opportunities to reconsider their initial votes. In section 8.5, interviewer assessments of various aspects of the interview are examined; and finally, section 8.6 presents a summary of our qualitative analysis.

See section 4.6.2.4.

§ 8.2 Examination of Responses to Open-Ended Questions

Several issues are considered in this section: whether respondents understood the choice; whether they took the choice seriously; whether they took relevant factors into account when they made their choice; whether they felt pressured to vote one way or another and, if so, whether there is evidence that this affected how they voted; and whether they were influenced by the presence of other people during the interview. The concern that underlies these issues is the meaningfulness of the respondents' voting choices, a concern that motivated both the NOAA Panel's methodological recommendations and the design and implementation of this study. Before addressing these issues, the method of coding the open-ended, verbatim responses into discrete-response categories is described.

§ 8.2.1 Coding of Open-Ended Questions

Periodically during the description of the injuries and the accelerated recovery program and immediately after the choice questions, the interviewers asked open-ended questions and recorded respondents' answers as completely as possible, word by word, in pen, on the questionnaire. The interviewers were also instructed to record in the same way spontaneous comments made by the respondent at any other time during the interview. The information recorded is referred to as a verbatim response.

The coding of the verbatim responses into discrete categories consisted of three steps. First, open-ended questions were selected for coding: A-7A, A-11A, A-13A, A-15A, W-1, W-4, W-5, and W-6. Second, after an examination of typical comments made in response to these questions in the Pilot IV interviews, coding schemes to categorize the various responses were developed. Next, two coders independently examined the open-ended and spontaneous verbatim

responses for the selected questions, separating them into individual ideas. Once this process was completed, the two coders worked together to negotiate resolutions of any disagreements they had.

In the third stage of coding, another set of two coders independently assigned each of the separate ideas into one of the categories listed in a preliminary coding scheme; after an initial batch of verbatims had been coded, we evaluated the discrepancies between the coders. We subsequently revised the coding scheme to increase precision and clarity. Each coder then independently assigned each of the verbatim ideas to one of the revised categories. Finally, they reconciled any codes on which they disagreed. This sort of procedure is conventionally used by psychologists and other social scientists to do content analysis of open-ended material (Bailey, 1987). The coded values are tabulated in Appendix D.2. and discussed below.

§ 8.2.2 Queries During Presentation of the Injuries and Accelerated Recovery Program

The first section of the interview contained four sets of questions (A-7/A-7A, A-11/A-11A, A-13/A-13A, and A-15/A-15A) that asked respondents if they wanted material repeated or if they had questions about material that had just been presented. Those who said *yes* to the first questions of the pairs were asked to describe what they would like repeated or what they wanted

---

See Appendix C.8 for a copy of the verbatim coding schemata.

Our pretesting indicated that some respondents wanted additional information about the sediment that would be used to cover the contaminated layer, but that the demand for this information was not broad enough to justify presenting these details to everyone and risking information overload. Therefore, we made this information optionally available to respondents by placing scripted responses in boxes at relevant places in the questionnaire (see Boxes 3, 4A, and 4B in Appendix A.1). The interviewers were instructed to read these scripted responses whenever a respondent mentioned these topics or asked questions about them and to record this action by checking the appropriate box.

to know. The responses to these questions provide useful information about respondents' reactions as the information on the injuries and accelerated recovery program was presented.

Question A-7 was asked after the description of the reproduction problems of the four affected species. When asked if there was "anything that I have told you about these four fish and bird species that you would like me to repeat?," 96 percent said *no*. Of the 62 respondents who said *yes*, and were asked in A-7A what they would like to have repeated, most inquired instead about aspects of the situation that would be described later. For example, 23 respondents wanted to know what had caused the four species' reproduction problems and 28 wanted information about these species' endangerment status. The survey instrument provided information about the cause of the reproduction problem and about endangerment in the very next section of the interview.

The second pair, A-11/A-11A, asked whether respondents had heard anything about the two chemicals that "are located in this particular place." The 146 respondents (8% of the sample) who answered *yes* were asked: "what have you heard?" Approximately 45 percent of these respondents made a clear reference to the DDT/PCB deposit off the South Coast.

Questions A-13/A-13A followed the description of the accelerated recovery program and asked the respondent if he or she had any questions about how it would work. Questions A-

---

In some cases, even though A-7 was coded *yes*, and a verbatim response was recorded at A-7A, the comment was clearly not in response to or relevant to the question. These types of comments were not considered in the coding of the A-7A verbatims nor were similar types of comments considered in the coding of A-11A, A-13A, or A-15A.

Respondents either asked questions about members of these species who lived elsewhere or whether the species may become extinct.

Here the interviewers were instructed to point to the location of the deposit on Map 3. See Appendix A.1.

15/A-15A followed the description of the natural recovery option and asked the respondent if he or she would like to know anything else about <u>either</u> the accelerated recovery program or the natural recovery option. Approximately 14 percent of all respondents asked questions at A-13A and at A-15A. Over 25 percent of respondents who asked a question at A-13A also asked a question at A-15A. The verbatim responses to these questions are summarized in Table 8.1.

Table 8.1 Verbatim Responses to A-13A and A-15A[a]

QUESTIONS ABOUT...	A-13A (N=257)	A-15A (N=260)
cost of program/paying for the speed-up program	28.4%	53.1%
how the speed-up program would work and its consequences	36.2%	19.2%
other possible ways to speed up recovery	16.3%	4.6%
natural recovery process	NA[b]	11.5%
whether the speed-up program would work	29.6%	18.1%
other	17.1%	14.6%

Percentaging base is the number of respondents who gave a verbatim response to each question. Percentag total more than 100 percent as multiple responses allowed.
This category was not applicable to A-13A.

The coded categories are of three types. As one might expect, the most commonly asked question concerned the cost of the program and, in particular, what the respondent would have to pay. Based on those who gave a verbatim response at A-13A and A-15A, 28 percent asked this type of question at A-13A and 53 percent at A-15A (the latter question asked just prior to the description of the program cost). The second type of query was about various aspects of how one or the other of the alternatives would work, including alternative ways to accelerate recovery.

The same coding categories were used for A-13A and A-15A except that a natural recovery category was added for the latter question.

The last type of query often involved expressions of skepticism about whether the accelerated program would actually work. Overall, the number of respondents who asked a question about either the injuries, accelerated recovery program, or natural recovery process at A-7, A-11, A-13, or A-15 was not large; and the questions they raised generally related to the material in a meaningful fashion.

§ 8.2.3 Did Respondents Take Relevant Factors Taken into Account When Voting?

To increase confidence that the voting choices are reliable, it is desirable that they be related to: (1) what the program would offer, (2) the cost of the program to the respondent's household, and (3) the respondents' preferences for environmental amenities of this sort. Important sources of evidence for these relationships are presented in subsequent chapters. These include the sensitivity of respondents to the size of the dollar amounts they would pay (Chapter 9), the construct validity equation (also discussed in Chapter 9), and the test of sensitivity to the scope of the injury (Chapter 10). Another source of evidence, particularly relevant to the first two items, is the set of respondents' answers to the open-ended, follow-up questions, asked immediately after the choice questions, which gave respondents the opportunity to explain why they made the choices they did. W-4 was asked of those who said they would not vote for the program, W-5 of those who said they were not sure about how they would vote, and W-6 of those who said they would vote for the program at either of the tax amounts they were asked about.

The NOAA Panel recommended the use of such questions and that their answers be carefully coded to show the types of responses (Arrow *et al*., 1993; p. 4609). They also noted that the open-ended responses should be explained by "making reference to the cost and/or the

value of the program". While we have made use of these types of questions in the Alaska survey (Carson *et al.,* 1992) and believe that they provide useful information in this study, in section 4.6.2.4, we called attention to the psychological literature on the reliability of introspective questions which suggests caution in interpreting these types of responses. There are two reasons why respondents' explanations will not necessarily be a complete accounting of all factors that shaped their judgements. First, a number of psychological studies suggest that people are sometimes unaware of factors that shape their own thinking and actions (*e.g.*, Nisbett and Wilson, 1977), and they sometimes forget about factors that influenced judgments made previously (Lodge, McGraw, and Stroh, 1989). Therefore, we expect some respondents may fail to mention considerations that shaped their voting decisions in this survey.

Second, and perhaps more important for this survey, is that in typical every-day conversations, speakers conform to certain norms or conventions, including the notion that one should not waste time telling someone else what that person already knows (Grice, 1975). In this survey, respondents likely recognized that the interviewers were well aware of all the details of the accelerated recovery program. Therefore, when explaining decisions to vote in favor of the program, respondents may have left out the specific factors that influenced their decisions. Rather, they may at times have simply made general, broad statements (*e.g.*, "the program will help the environment") that were intended to summarize what they have been told but in different words and without being unnecessarily redundant. For these reasons, we expected respondents' answers to the follow-up questions to provide insight into, though not necessarily a complete accounting of, the factors influencing their choices.

W-4, administered to respondents who voted *against* the program at both the first and second choice questions (W-1 and W-3), asked: "Did you vote against the program because it

isn't worth <u>that</u> much money to you, <u>or</u> because it would be somewhat difficult for your household to pay that much, or because of some <u>other</u> reason?" These particular response categories were offered in the question itself to ease the possible embarrassment some respondents may have felt about saying that they couldn't afford the tax amount asked about. Overall, 15.6 percent of those who answered this question chose the "somewhat difficult to pay" response. The likelihood of giving this response was strongly related ($p < 0.001$) to the tax amount the respondents were asked about in W-1, with twice the percent of respondents at the higher amounts saying they could not afford it than the percent at the lower amounts.

A little over 12 percent gave the first-offered response, "isn't worth that much money", and 74 percent gave a different reason. If the respondent said he or she had another reason, the interviewer was instructed to probe to learn what that reason was. Coders assigned each of the reasons expressed in these "other" verbatim responses into the categories shown in Table 8.2. In order to give a complete picture of the responses, also included in this table (shown in italics) are the answers to the two pre-coded categories.

The most common type of response was the view, held by 51.5 percent of those who answered W-4, that the problem described in the scenario was not that important and/or other problems are more important to them. Respondents who expressed this view mentioned reasons like: the reproduction problems will eventually take care of themselves; the injury is just in one area; and other types of problems concern them more such as the homeless, schools, and crime. As noted above, another 12.5 percent chose the related pre-coded response, "the program isn't

---

The format of this question is identical to the comparable question in the Alaska survey (Carson *et al.,* 1992, Question A-18).

In 15 cases, the interviewer circled more than one W-4 answer category, hence the percentages total more than 100.

worth that much money".

Table 8.2 Reasons for Choosing to Vote *Against* the Program

W-4. Did you vote against the program because it isn't worth <u>that</u> much money to you, <u>or</u> because it would be somewhat difficult for your household to pay that much, or because of some <u>other</u> reason?	
CODING CATEGORY	PERCENTAGE [N=825][a]
Problem not that important/Other problems more important	51.5%
Somewhat difficult to pay/Cost too high	26.3%
Concerns about program or payment plan design	21.2%
Isn't worth that much money	12.5%
Wants more information	2.1%
Other	9.3%

Percentaging base is the number of respondents who answered W-4 and/or gave a response "other (specify)". Categories in italics were assigned by the interviewers (*i.e.,* pre-c answer categories). Percentages total more than 100 percent as multiple responses allowed.

Overall, a third of the respondents mentioned some aspect of the program cost. Twenty-one percent mentioned various concerns they had about the program, such as skepticism about whether it would work or whether the State would really use the money for the stated purpose. (Forty percent of these respondents also gave reasons that involved the cost or the relative unimportance of the program.) The W-4 responses displayed in Table 8.2 and our further analysis of these responses strongly suggest that respondents who voted *against* the program were

The overlap between these two categories is less than 1%.

This includes respondents who spontaneously mentioned some aspect of cost when they were first asked the W-1 question where the cost of the program was first revealed.

attentive to the object of choice and to the financial implications of voting for it and that they weighed the object of choice against other concerns when making their decision.

Respondents who said at W-1 that they would not vote for the program or were not sure about how they would vote at W-1 and who, in addition, indicated at W-3 that they were *not sure* about how they would vote, were asked W-5: "Could you tell me why you aren't sure?" As shown in Table 8.3, the verbatim responses given by the 99 respondents who were asked W-5 are similar to the reasons respondents gave for voting against the program in W-4. A strong plurality commented that the problem was not that important or other problems were more important to them. Twenty-seven percent said the cost was too high. As one might expect, those in the *unsure* category were more likely than those who voted *against* to mention concerns about the program or the design of the payment plan. They were also much more likely to express a desire for more information. However, only about 2 percent of the total sample mentioned lack of information as a reason for why they were not sure. Thus, it appears that the information provided in the interview was sufficient for most respondents to make a choice.

Table 8.3 Reasons Why *Not Sure* About Program Vote

W-5 Could you tell me why you aren't sure?	
CODING CATEGORY	PERCENTAGE [N=99][a]
Problem not that important/Other problems more important	41.4%
Cost too high	27.3%
Concerns about program or payment plan design	31.3%
Wants more information	23.2%

The total of those who gave this reason at W-4 or W-5 divided by 1857, the base sample size.

Other	30.3%

Percentaging base is the number of respondents who gave a response to W-5. Percenta
total more than 100 percent as multiple responses allowed.

Question W-6, administered to those who voted for the program at *either* W-1 or W-3, asked: "People have <u>different</u> reasons for voting for the program. Can you tell me what covering the contaminated sediments would <u>do</u> that made <u>you</u> willing to pay for it?" This wording, which is similar to the wording used for the comparable question in the Alaska study, was designed to help overcome the conversational convention—that one should not tell someone what they already know—by focusing the respondent on the outcome of the program. The interviewers were trained to use neutral and nondirective probes when respondents gave answers that seemed vague or non-responsive to the question to determine whether the respondent had anything more specific in mind.

The W-6 verbatims were coded into the categories listed in Table 8.4. The percentage distribution across the categories for the 907 respondents who answered this question shows that a large majority, 71.9 percent, voted for the program to help the affected species or area by covering up the contaminated sediment. Twenty-two percent mentioned that hastening the recovery process was important to them because they did not want to wait for natural recovery to take place. The third most common (16%) type of reason was expressions of personal interest in the program because it would realize goals that were important to the respondent. These reasons were prefaced by "I" or "we" (*e.g.,* "I am really concerned about those two birds [eagle and falcon]" or "we like the natural environment around us"). Other reasons in this category reflected

---

Carson *et al*., 1992, Question A-20.

As noted in Chapter 6, a chapter of the interviewer's manual for this study is devoted to probing

the respondents' personal interest in water-based recreation activities such as fishing.

Table 8.4 Reasons for Choosing to Vote *For* the Program

W-6. Can you tell me what covering the contaminated sediments would <u>do</u> that made <u>you</u> willing to pay for it?	
CODING CATEGORY	PERCENTAGE [N=907][a]
Help affected species and/or the area where they live	71.9%
Hasten the recovery process	21.7%
Respondent personally concerned about environment/wildlife or perceives household would benefit in some way	16.4%
Prevent possible physical harm to respondent or others	13.9%
Feel responsible to help fix this problem	13.6%
Others such as grandchildren or people living in the area would benefit	12.0%
Cost affordable/reasonable	9.8%
Might help other animals/ecosystem	7.5%
Protect environment[b]	1.9%
Other	15.6%

Percentaging base is the number of respondents who gave a response to this question. Percer total more than 100 percent as multiple responses allowed.

Only includes those for whom no other category was coded.

The reasons coded in the category "prevent possible physical harm to respondent or others" usually involved a desire to avoid the possibility of having to worry, for themselves or others, about catching or eating contaminated fish. We had anticipated that some respondents would continue to be concerned about this despite the assurances they were given during the interview that the fishing ban would prevent harm to humans. This is why the first reconsideration question, W-7, addressed this concern and highlighted that the only outcome of

(Westat, 1994a).

the program would be the accelerated recovery of the four species. Those who had expressed a concern about possible physical harm to humans at W-6 were significantly more likely to change their vote at W-7 compared to other types of respondents (p = 0.008). This is reassuring evidence that respondents paid attention to the information conveyed in W-7. Furthermore, none of the spontaneous remarks made at W-7 by respondents who reaffirmed their willingness to vote for the program indicated that they were unwilling to accept the question's premise that only the four species would be helped by the program.

Among the other types of reasons were expressions of personal or collective responsibility to do something about the problem because it was caused by humans (13.6%), and satisfaction that others, such as grandchildren or people living in the area, would benefit from the accelerated recovery (12%). About 10% of the respondents mentioned that the cost was reasonable given what the program would accomplish. Next were those who mentioned that it might help other animals (7.5%).

The most commonly expressed concern was that the presence of the chemicals in the local ecosystem could also affect other, unspecified, creatures. Those who are shown in Table 8.4 as giving a response coded as "protect environment" (1.9%) are those who did *not* clarify this thought by giving any other type of reason in their answer to W-6. Other respondents giving answers coded in this category, as expected, appeared to use this type of comment to refer to what the respondent presumed the interviewer already knew and clarified this with more specific reasons in response to the non-directive probes.

Other characteristics of those who reconsidered their *for* votes are discussed in section 8.4.

Eighty respondents changed their vote at W-7; see section 8.4 for a more detailed discussion.

The total percent who originally gave reasons (*i.e.*, response before interviewer probed) coded in

Verbatim responses to W-4, W-5, and W-6 suggest that respondents took relevant factors of cost and the value of the program to them into account when making their voting choices. Moreover, there is very little evidence that respondents who voted for the program did so because they were misinformed about what the program would and would not accomplish. Those who gave a "prevent possible physical harm" answer were more likely to change their vote from *for* to *against* when reminded at W-7 that the human health was not affected. Moreover, only 9 respondents who gave an answer related to possible physical harm did not also give another reason for why they voted *for* the program. Finally, the percentage giving reasons coded as "might help other animals/ecosystem" is small (7.5%), and only four of these respondents gave just this reason. Few of these respondents mentioned the names of other animals they had in mind; instead, they referred to the likelihood the local ecosystem might be generally affected.

§ 8.2.4 Did Respondents Feel Pressured to Vote One Way or Another?

Question B-7 asked respondents whether they perceived that the interview, overall, tried to push them to vote one way or another or let them make up their own mind. Seven percent of the total sample, or 132 respondents, said that they thought the interview had tried to push them or were not sure about this. B-7A asked these respondents: "<u>which</u> way did you think it pushed you?" Of the 132 respondents who answered this question, 101 (5.4% of the total sample) felt pushed to vote *for* the program, 26 (or 1.4%) felt pushed to vote *against*, and 5 respondents were not sure about the direction. All were asked to explain in B-7B: "What was it that made you think that?" Some of those who said they felt pushed to vote *for* had no specific reason in mind, just a generalized "feel" about this. Others in this category mentioned the fact of being presented

the "protect environment" category was 17.4 percent.

with all the information about the injuries or being asked the follow-up choice question (W-2/W-3).   Those who said they felt pushed to vote *against* were likely to mention the positive information that the species would recover on their own or Card H (which listed reasons why one might vote against the program).

Table 8.5 shows the relationship between the perceived direction these respondents felt they were pushed and how they had voted at W-1. Those who felt pushed to vote *for* voted

Table 8.5 Voting Patterns by Direction Felt Pushed

Direction Felt Pushed	Voted *For*	Voted *Not For*
Pushed *For* [N=101]	43.6%	56.4%
Not Pushed [N=1707]	40.8%	59.2%
$\chi^2_{(1)} = 0.31; p = 0.579$		

Direction Felt Pushed	Voted *For*	Voted *Not For*
Pushed *Against* [N=26]	19.2%	80.8%
Not Pushed [N=1707]	40.8%	59.2%
$\chi^2_{(1)} = 4.93; p = 0.026$		

for the program with virtually the same frequency (p=0.579) as the 92 percent of the sample who said they felt the interview let them make up their own mind.   This is consistent with the interpretation that although they believed they may have felt some pressure, they did not seem to be influenced by it. In contrast, there is a significant difference (p=0.026) between those who felt pushed to vote *against* the program and the rest of the sample, with those who felt pushed to vote *against*, voting against more often than those who felt the interview let them make up their own

---

See Appendix A.1.

Two respondents changed their vote after B-7 was asked. See Table 8.8 in section 8.4.

mind.

§ 8.2.5 Were Respondents' Choices Influenced by Others?

In order to avoid distractions, interviewers were instructed to refrain from conducting interviews with other persons present. However, in a number of cases, living arrangements were such that someone else was present during some or all of the interview. Frequently, these were young children in the respondent's care. In order to differentiate these cases from those where teenagers or adults were present, the interviewers were asked in D-6 to report whether anyone age 13 or older was present when the respondent voted. The answer was positive in 22 percent of the interviews. Judging from interviewer remarks recorded on the questionnaires, almost all of these individuals were other household members.

In question D-6A, the interviewer was asked whether he/she thought the other person(s) affected how the respondent voted. In almost 90 percent of the cases where someone age 13 or older was present while the respondent voted, the interviewers judged that there was no effect. There were 15 cases (less than one percent of the total sample) where the interviewer said he/she believed that the other person present did have an effect and 26 cases where the interviewer indicated that he or she did not know. We examined the D-6a and D-7 verbatim comments for these cases. Whenever influence was mentioned, it was almost always by another household member.

§ 8.3 Section B Debriefing Questions

Respondents were asked to make a choice between a program to accelerate recovery, which would occur in five years and cost their household a specified amount in higher taxes, and

natural recovery, which would occur in fifty years and not cost their household anything more in higher taxes. As the NOAA Panel pointed out, the reliability of respondents' choices depends on the degree to which they accepted or believed certain basic assumptions underlying the choice. For example, to the extent that some respondents did not believe that the accelerated recovery program would be effective, their choices would tend to under-represent their value for accelerating recovery. This is because they believed the program would be less helpful in accelerating recovery than we had intended them to believe. The reverse would be the case if some respondents believed that natural recovery would take longer than the 50 years. In this case, their choices would be based on the assumption that a longer than intended stream of benefits would be created if the program were implemented. As the NOAA Panel commented (with reference to what happens when respondents do not accept information of this type): "in effect they (the respondents) will be answering a different question from that being asked." (Arrow, p. 4605).

During our research, we devoted a great deal of effort to developing a program that would be perceived by as many respondents as possible to be both effective in accelerating recovery and targeted to the specific injuries. The presentation of the natural recovery option received a similar amount of attention to also make it as credible as possible. As we will show in this and following chapters, the available evidence indicates that we were quite successful in this regard. Nevertheless, given the diversity of respondent experiences and levels of trust in information they receive from the government, the choice perceived by *some* respondents differed somewhat from the one that was described to them.

Mitchell and Carson (1989; pp. 249-252) discuss this issue at length.

The effect that lack of acceptance has on estimates of WTP is investigated in Chapter 9; the

The data examined here are from a series of questions asked at the beginning of Section B of the survey to check on respondent acceptance of several elements of the choice, including two key items, the length of time that natural recovery would take, and the effectiveness of the program to accelerate recovery. These questions asked respondents what they had in mind about these choice elements when they voted. As noted in Chapter 6, this type of introspective assessment may be unfamiliar to respondents, so these questions were carefully designed to avoid misunderstandings (*e.g.,* respondents taking them as an invitation to speculate about the topic of the question instead of reporting what they had been thinking at the time they decided how to vote). The wording we finally adopted appeared to have overcome most of these problems.

§ 8.3.1 DDT/PCB's and Reproduction Problems

The first debriefing question, B-1, asked if it seemed to the respondent that "DDT and PCB's could cause the reproduction problems I told you about." Almost all the respondents accepted this basic premise, with 94 percent answering *yes*. Those who said *no* (2.7%) or *not sure* (3.7%) were disproportionately likely to be among those voting *not for* the program (p < 0.001).

§ 8.3.2 Length of Natural Recovery

---

effect tends to lower our estimates.

  As noted in Chapter 6, the interviewers were told that "sometimes respondents wonder why they are being asked questions like this about what they were thinking when they answered the vote questions — after all, they might say, you told them that it would take fifty years, why should they doubt it." The interviewers were instructed to tell such respondents: "We find that some people have different ideas about this. It is important for us to know what <u>you</u> had in mind" (Westat, 1994a; p. 4.71.).

The next question, B-2, asked about a key feature of the natural recovery: how long respondents had assumed it would take. Seventy percent said they had assumed that it would take about 50 years when they decided how to vote. Because we were interested in whether their beliefs differed significantly from this time frame, those who said *no* or *not sure* in response to B-2 were asked in a follow-up question, B-3: "Did it seem to you that it would take a lot more than 50 years or a lot less than 50?" Table 8.6 summarizes the B-2/B-3 responses. Some respondents (6%) said they assumed recovery would take a lot more than 50 years. Others (15%) said they believed recovery would take a lot less time than this. About 9 percent expressed other views, which consisted mainly of expressions that no one could know for sure about the time frame or the belief that it would take just a little more or a little less than the 50 years. Those who felt natural recovery would take a lot <u>more</u> than 50 years were significantly more likely to vote *for* the program (p < 0.001); and those who felt natural recovery would take a lot <u>less</u> than 50 years were significantly less likely to vote *for* the program (p < 0.001).

Table 8.6 Respondents' Assumptions About Length of Natural Recovery[a]

COMBINED RESPONSES TO B2/B3	PERCENTAGE [N=1849]
A lot more than 50 years	6.3%
About 50 years	69.8%
A lot less than 50 years	15.0%
Other/Not sure[b]	8.9%
TOTAL	100%

Percentaging base is the number of respondents who answered B-2
Those who said "other" or "not sure" to B-3.

Seventeen of these "other" views clearly indicated a direction of the divergence from the 50 year natural recovery period and were recoded into the *lot more* or *lot less* categories.

§ 8.3.3 Effectiveness of Accelerated Recovery Program

Another key respondent assumption examined in section B was how effective the respondents believed the accelerated recovery program would be in solving the reproduction problem within five years. Question B-4 asked: "<u>When</u> you decided how to vote, did it seem to you that the speed-up program would be <u>completely</u> effective in solving the reproduction programs within five years?" Those who said "no" or "not sure" in response to this question were asked in B-5 which of four degrees of effectiveness they thought the program would accomplish. Table 8.7 summarizes the B-4/B-5 responses.

Table 8.7 Respondents' Perceptions About Effectiveness of Program

COMBINED RESPONSES TO B-4/B-5	PERCENTAGE[N=1848][a]
Completely effective	52.4%
Mostly effective	11.0%
Somewhat effective	24.7%
Not too effective	6.2%
Not effective at all	2.6%
Not sure	3.1%
TOTAL	100%

Percentaging base is the number of respondents who answered B-4

As shown in the table, 52% percent indicated, that when voting, they thought the program would be completely effective. Another 11 percent of the sample thought the program would be "mostly effective" and a quarter said "somewhat effective". Only nine percent held serious doubts about its effectiveness (answering either "not too effective" or "not effective at all") and an additional 3 percent expressed uncertainty about its effectiveness. Given potential respondent concerns about the possible effects of earthquakes or ocean currents on the deposit and their general skepticism about government promises, this level of acceptance is reassuring. As shown in Chapter 9, respondents who did not think the program would be completely or mostly effective were less likely to vote *for* the program.

Further insight into the effects of nonacceptance can be gained by looking at nonacceptances at both B-1 and B-5. A total of 198 respondents (10.7% of the sample) didn't accept that DDT and PCB's could cause the injury (B-1) and/or believed that the accelerated recovery would be "not too" or "not" effective or both. Only a few of these respondents (14 of 198) chose to vote *for* the program.

§ 8.3.4 Length of Payment

Question B-6 asked respondents whether they thought their households would have to pay the special tax for the program "for one year or for more than one year?" Sixty-three percent said one year, while 28 percent said they had doubted that it would be just one year when they voted. This level of skepticism about the promise that the State would only require a one-time payment reflects the frequently cynical views expressed by participants in focus groups and in pretesting. Here again, as will be shown in Chapter 9, this lack of acceptance is associated with a lower

willingness to pay for the program as one might expect if respondents believed the object of choice actually entailed a higher cost than was described to them.

§ 8.4 Reconsideration of *For* Votes

Three questions included in the survey instrument gave respondents an opportunity to change their votes.   In each case, only respondents who voted *for* were offered these opportunities.   The first reconsideration opportunity was presented in W-7, which appeared immediately after the choice questions.   Respondents were told to "suppose human health is definitely not affected" and the program would only help the four species of fish and birds. W-7 then asked: "Would you vote for or against the program if it cost your household [the highest amount the respondent had voted for]?" While this question focused on the human health issue, it also offered respondents who wanted to reconsider their vote for other reasons an opportunity to do so.

The other two reconsideration questions, C-18 and C-20, were asked much later in the interview of only respondents who had voted *for* the program (and who had not changed their vote at W-7). They were based on certain types of answers to preceding "filter" questions. First, all respondents who voted *for* the program at W-7 were asked in question C-17 how difficult it would be for their households to pay that amount next year if the program passed.   Most respondents said it would not be "difficult at all" (36.7%) or "not too difficult" (33.3%) for them

---

As noted in Chapter 6, to help the interviewers keep track of how people voted and the highest amount voted for, the instrument had a fold-out skip record where the interviewer recorded how the respondent voted on W-1 to W-3. They also recorded the response at W-7 on the skip record so they could refer back to it at Box 8.

The response to question C-17 is strongly associated ($p < 0.001$) in the expected way with the amount asked in W-1.

to pay. Twenty-one percent said it would be "somewhat difficult" and 7 percent said it would be "very difficult." The 234 respondents who gave the last two responses as well as the 13 who were unsure at C-17 were given an opportunity to reconsider their vote at C-18.

Those who indicated that paying for the program would <u>not</u> be *somewhat* or *very difficult* for their household were asked, in a second filter question (C-19), how strongly they favored the program at the highest tax amount that they had previously voted for. Those who said they favored the program "not at all strongly" (N=8) or "not too strongly" (N=101) were given an opportunity to reconsider their vote at C-20. To be conservative, we counted those who said *not sure* to any of the three reconsideration questions (W-7, C-18, or C-20) as having changed their vote to *against*. These combined categories (*i.e., against* or *not sure*) are referred to as *not for* the program below.

Table 8.8 summarizes the reconsideration results. A total of 105 people, 11.5 percent of those who originally voted *for* the program, changed their votes from *for* to *not for*. Most respondents (N=80) who changed their vote did so at the first opportunity offered, W-7. One respondent volunteered at question C-21 (which asked about how strongly the respondent favored the program) that he or she no longer favored the plan at all so he or she was counted as having changed his/her vote.

Those who had just been given the chance to change their vote in C-18 were also asked the same strength question at C-21 if they did not reconsider their vote at C-18. Combining the responses to both of these identically worded questions shows that most of the respondents who chose to vote *for* the program favored the program at that tax amount "very strongly" (25%) or "strongly" (55%), while about 18% favored it "not too strongly" and just one percent "not at all strongly."

This is consistent with how we treat those who said *not sure* to the original voting questions in Chapter 9.

Table 8.8 Respondents Who Changed Vote from *For* to *Not For*

AT QUESTION ...	AGAINST	NOT SURE
W-7. Suppose human health is definitely not affected in this situation and the program would only speed up the recovery of these four species of fish and birds. Would you vote for or against the program if it cost your household $(highest amount respondent voted for)?	64	16
C-18. Now that you have had a chance to think a bit more about this, would you vote "For" or "Against" the program if it cost your household $(highest tax amount respondent voted for)?	8	13
C-20. *Same wording as C-18.*	2	1
C-19/C-21 "Doesn't favor plan"	1	0

There are various distinguishing characteristics of those respondents who changed their votes. Respondents in the lowest three income categories were almost twice as likely to change (p=0.003) as other respondents providing initial *for* votes. Other categories of respondents who were more likely to switch include those who in A-2E favored reduced spending on protecting endangered wildlife (p=0.002), those who in B-4/B-5 thought that the plan would not be completely or mostly effective (p=0.003), and those who thought that natural recovery would take much less than 50 years (p=0.003). Similar significant patterns of an increased propensity to change were observed among households which did not engage in saltwater recreation, bird watching, or watching television nature shows.

§ 8.5 Interviewer Evaluations

Another source of information about whether respondents understood the voting choice is the series of questions in section D which the interviewers answered shortly after completing the

interview. The items listed in D-1 asked the interviewers to assess the respondent's reactions "as you read through the material beginning with A-3 through A-16." This is the portion of the interview that presented the elements of the choice, such as the nature of the injuries, their cause, the accelerated recovery program, and the natural recovery process. Table 8.9 shows the interviewer ratings for how well the respondent understood this material and also for how distracted and attentive the respondent was during the presentation. The interviewers rated 28 percent of the respondents as understanding this material "extremely" well and 59 percent "very" well for a total of 87 percent in these two categories. Only one percent were rated as understanding it only "slightly" or "not at all," and the remaining 12 percent as understanding it "somewhat." Very low percentages of respondents were said to be "extremely" or "very" distracted (2%) during the presentation and/or "slightly" or "not at all" attentive (1%) to this material.

Table 8.9 Interviewer Evaluation of Respondent Reaction to Choice Elements

D-1 Items	Extremely	Very	Somewhat	Slightly	Not at all	Not Sure
How distracted was the R?	0.3%	1.6%	8.6%	19.9%	69.1%	0.1%
How attentive?	27.0%	58.5%	12.8%	1.0%	0.3%	0.0%
How well did the R. understand this material?	27.5%	59.0%	11.8%	1.0%	0.1%	0.3%

Question D-2 asked if the respondent had said anything that suggested he or she had any difficulty understanding either the accelerated recovery program or the natural recovery process. A total of 46 respondents (or 2.5 percent of the total sample) were identified as having had a

difficulty of some sort. Of those said to have had a difficulty, only 13 gave a *final for* vote. The

interviewers were asked in an open-ended question, D-2A, to "describe the difficulties". From the

interviewers' descriptions, many of the difficulties appeared to be overcome to the interviewer's

satisfaction.

Other section D questions asked for the interviewer's impression of the respondent's

reaction to the choice questions (W-1 through W-3). D-3 asked if the respondent had *any*

difficulty understanding them and, if so, to describe the difficulties (D-3A). Thirty-nine

respondents (2% of the total sample) were identified in this category, of whom 13 were *final for*

voters at the W1AMT asked about. The difficulties described by the interviewers for these 13

respondents included hearing problems, minor misunderstandings that were subsequently clarified,

or reiterating respondent questions about aspects of the program.

On the basis of the set of interviewer evaluation questions we have just explained—D-1A,

D-1B, D-1C, D-2, and D-3—there are 120 respondents, representing 6.5 percent of the sample,

who may have had a problem understanding or responding to the choice questions. This is a

rather inclusive measure because, as we mentioned, some of these people may not have had a

problem. For example, the mere fact of being identified by the interviewer in D-2 as having

difficulty understanding the injuries or program, does not necessarily mean that the difficulty

interfered with a respondent's ability to make a meaningful choice. Do these cases contribute to

---

In what follows, references to *final for* votes refer to those who did not revise their original vote
for the program at a later point in the interview. Those identified as having difficulty at D-2 were
more likely (p< 0.001) to change their vote.

We define this inclusive variable, PINTPROB, as being equal to 1 if ([D-1A=1 or D-1A=2;
"extremely" or "very" distracted] or [D-1B=4 or D-1B=5; "slightly" or "not at all" attentive] or
[D-1C=4 or D-1C=5; "slightly" or "not at all" understand injuries and program] or [D-2=1;
respondent indicated difficulty understanding injuries or program] or [D-3=1, respondent
indicated difficulties understanding vote questions]) and 0 otherwise.

an overestimate of our estimate? An examination of the data suggests that they do not. First, they represent a relatively small fraction of the sample. Second, they are much more likely to change an initial *for* to a *not for* vote during the three reconsideration opportunities (p<0.001). Third, as we will see in section 9.5, after taking the reconsidered answers into account, the amount of money this group of respondents is willing to pay for the program is substantially <u>lower</u> than the rest of the sample.

Another factor that might affect a respondent's understanding of the choice is whether he or she was impatient to get through the interview. Questions D-4 and D-4A asked the interviewer to rate the degree of impatience the respondent had when he or she was asked the voting questions. The vast majority of the respondents (83%) were not thought to be impatient, and another 8 percent were rated as "not very" impatient or only "a little" impatient, for a total of 91 percent. Five percent were said to be "somewhat" impatient, and 3 percent said "very" impatient.

Interviewer ratings of "how serious was the consideration the respondent gave to the decision about how to vote" (D-5) can be used to examine another goal of this study: to develop a plausible choice mechanism which the respondents would take seriously. As shown in Table 8.10, 81 percent of the total sample were thought to have given the matter "very" or "extremely" serious consideration. Only about 2 percent or 41 cases were rated as giving it only "slightly" or "not at all" serious consideration. These respondents were somewhat less likely to give a *final for* vote; however, although this difference is suggestive, it is not quite statistically significant (p=0.102)

Table 8.10 Interviewer Evaluation of the Seriousness of Respondent Consideration of the Voting Decisions

Question D-5	Extremely	Very	Somewhat	Slightly	Not at all	Not Sure
How serious was the consideration the R. gave to the decision about how to vote?	25.0%	55.8%	16.0%	1.8%	0.4%	0.4%

§ 8.6 Summary

The pattern of responses to the various open-ended questions we considered in this chapter were consistent with those one would expect if respondents were paying attention to the material and evaluating the object of choice as intended. The answers to the questions about why they made their voting choices (W-4 to W-6) referred to relevant features of the accelerated recovery program such as its cost and what the program would accomplish. These answers help provide insight into the reliability of the voting choices, one of the topics also examined in Chapter 9. The respondents' reasons for their choices are not used in that chapter's quantitative analysis, however, as they are too closely associated with the choice variable used as the dependent variable in the multivariate choice function.

The debriefing questions in Section B of the survey obtained information about the degree to which respondents accepted various features of the injuries and the accelerated recovery program. These included the role of DDT and PCB's in causing the injuries, the length of natural recovery, the effectiveness of the accelerated recovery program, and the duration of the special tax for the program. Overall, the number of respondents who did not accept the scenario, such as not believing that the DDT and PCB's could cause the injuries, or that the accelerated recovery program would not be effective, is small. These respondents, as well as those who thought they would have to pay for the program for more than one year, are less likely than the rest of the

sample to vote *for* the program; hence, the resulting effect of a lack of acceptance of these features tends to decrease willingness to pay. These issues are further examined in sections 9.5 and 9.6 of Chapter 9.

An important feature of our design was to offer the respondents who voted *for* the program opportunities to reconsider their choices. This was done both shortly after they voted (W-7), and later after they had more time to consider the implications of their choices (C-17 to C-21). Those who gave a W-6 verbatim response related to "possible physical harm" as well as the small number of respondents who the interviewers identified as potentially problematic were more likely to reconsider and change their *for* vote to an *against* vote than the rest of the sample. The principal measure of respondent choices used for the analysis in Chapter 9 is based on these final choices.

The principal finding of our analysis of the interviewer debriefing questions in Section D of the survey was that there was very few cases where the interviewers identify possible problems with respondents' attentiveness, comprehension, and impatience. Those who the interviewers identified as problematic had a substantially lower willingness to pay than the rest of the sample. The interviewer debriefing questions are used in a more detailed analysis in section 9.5 of Chapter 9.

§ 9 Analysis of Choice Questions

§ 9.1 Introduction

This chapter presents a lower-bound estimate of prospective interim lost use value (ILUV) constructed from respondents' choices in the base survey instrument and examines the relationship between those choices and other variables measured by the survey. In section 9.2, two choice measures are summarized, one based on the responses to the W-1 choice question and the other based on the adjusted responses to that question after respondents were given opportunities to reconsider their vote *for* the accelerated recovery program. Section 9.3 introduces the non-parametric (Turnbull, 1976) statistical framework used in our analysis and discusses the statistical properties associated with measures of central tendency for the willingness-to-pay (WTP) distribution. Section 9.4 provides an estimate of the lower-bound mean value for the sample.

In section 9.5, bivariate relationships between the choice measures and other variables measured by the survey are examined. Cross-tabulations of the primary choice measures with specific types of variables recommended by the NOAA Panel are included in this section. In section 9.6, construct validity is examined using a multivariate counterpart to the evaluations reported by individual variables in section 9.5. Section 9.7 examines the implications of setting all respondents who said that they did not pay California income taxes to *against* program votes. Finally, in section 9.8, the results of the analysis are summarized.

Chapter 10 compares the choices made in the base and scope surveys, and Chapter 11 applies population weights to the choice data analyzed in this chapter to arrive at our estimate of prospective ILUV.

§ 9.2 Defining Choice Measures

The principal choice question in the survey was W-1, which asked respondents if they would vote for or against the accelerated recovery plan if it cost their household a pre-assigned tax amount. Respondents were randomly assigned to one of five different W-1 tax amounts: $10, $25, $80, $140, or $215. This W-1 tax amount will be referred to as W1AMT. Responses to the W-1 choice question by W1AMT are shown in Table 9.1a.

Table 9.1a W-1 Response by W1AMT

W1AMT	For	Against	Not Sure
$10	59.4%	35.3%	5.3%
$25	51.4%	42.6%	6.0%
$80	37.0%	54.5%	8.5%
$140	31.7%	60.8%	7.5%
$215	24.7%	68.8%	6.6%

In the analysis that follows, the *against* and *not sure* categories (displayed in the last two columns of Table 9.1a) are combined into a single *not for* category; this coding is referred to as the **W1** choice measure. Table 9.1b displays the percentages of *for* and *not for* responses to **W1** by W1AMT. Based on these percentages, a $\chi^2_{(4)}$ test (126.39) clearly rejects the null hypothesis (p < 0.001) that the percent *for* does not systematically vary with W1AMT.

Table 9.1b W1 Choice Measure by W1AMT

W1	For	Not

The sample marginal distributions for the discrete response questions in the base survey instrument are provided in Appendix D.1.

All choice measure variables are denoted in bold capital letters.

AMT			For	
$10		59.4 %		40.6 %
$25		51.4 %		48.6 %
$80		37.0 %		63.0 %
$140		31.7 %		68.3 %
$215		24.7 %		75.3 %
$\chi^2_{(4)} = 126.39$; $p < 0.001$				

A choice measure defined only by W-1 responses (*e.g.,* the **W1** choice measure defined above) results in what is referred to as *single-bounded* interval data. That is, if a respondent votes *for*, we know that the respondent's willingness to pay for the program is bounded from below by W1AMT (*i.e.,* the respondent is willing to pay at least W1AMT.) If the respondent gives a *not for* answer, we assume that the respondent's willingness to pay is bounded from above by W1AMT (*i.e.,* the respondent may be willing to pay some tax amount below W1AMT or may not be willing to pay anything at all).

Respondents who voted *for* the program at W-1 were then asked about a pre-assigned, higher tax amount (W2AMT) in the follow-up choice question, W-2; those who voted *not for* (*i.e.,* voted *against* or were not sure about their vote) at W-1 were asked about a pre-assigned, lower tax amount (W3AMT) in the follow-up choice question, W-3. The five sets of tax amounts used in the W-1, W-2, and W-3 questions are shown in Table 9.2.

Table 9.2 Tax Amount by Version and Choice Question

Version	W-1	W-2	W-3

1	$10	$25	$5
2	$25	$45	$10
3	$80	$140	$45
4	$140	$215	$80
5	$215	$360	$140

Combining responses from W-1, W-2, and W-3 results in what is often referred to as *double-bounded* interval data (Hanemann, Loomis, and Kanninen, 1991). This term is used because combining the response to the first question with that to the second question locates the respondent's WTP in a tighter interval (*i.e.*, below W3AMT, between W3AMT and W1AMT, between W1AMT and W2AMT, or above W2AMT) rather than simply above or below W1AMT. We will refer to the choice measure based on the combined responses to W1, W2, and W3 as **WDB**.

In this chapter, the results based on the single-bounded interval data are presented. As the analysis of the double-bounded data yields similar conclusions to that based on the single-bounded data, the double-bounded results are presented only in footnotes and in Appendix F tables.

Respondents who voted *for* the program at either W-1 or W-3 were offered opportunities to change their vote. The first opportunity to do so was in question W-7, and the second in the C-17 to C-21 question sequence administered in the final section of the interview. Revising the **W1** choice measure (defined above) to take into account those respondents who reconsidered

---

This approach ignores the bias, typically downward, that theory suggests may be present in the second response. This bias may occur due to strategic incentives to misrepresent preferences introduced by the second question and because the second question may change the perceived characteristics of the object of choice.

See sections 6.4, 6.8, and 8.4.

their *for* vote results in a second choice measure; this choice measure is referred to as **W1CH**. Because only respondents who voted *for* the program were given an opportunity to change their votes, **W1CH** is, by construction, a more conservative choice measure than **W1**.

Table 9.3 displays the **W1CH** choice measure by W1AMT. Based on these percentages, a $\chi^2_{(4)}$ test (126.93) rejects the null hypothesis (p < 0.001) that the **W1CH** choice measure does not systematically vary with W1AMT. The **W1CH** choice measure is used for most of our analysis in this chapter (and the following chapter) as it represents the respondents' final choice.

Table 9.3 W1CH Choice Measure by W1AMT

W1AMT	For	Not For
$10	55.9%	44.1%
$25	46.3%	53.7%
$80	32.9%	67.1%
$140	26.5%	73.5%
$215	22.3%	77.7%
$\chi^2_{(4)} = 126.93$; p < 0.001		

§ 9.3 Statistical Framework for Analysis

The final selection of a summary statistic is always a professional judgment that reflects the relative importance of different properties of the estimator given the goals underlying the

---

The respondent's last *for* response to W-1, W-2, or W-3 was modified from a *for* to *not for* based on his/her answer to W-7 and the relevant components of the C-17 to C-21 question sequence. The small number of respondents who voted *for* to the W2AMT (asked about in W-2) but later reconsidered their vote, changing it to a vote *not for*, were treated as *not for* votes in constructing the **W1CH** choice measure, even though it is possible that some of these respondents would still have been willing to pay the W1AMT tax amount.

See Appendix E for a more technical description.

analysis. Our objective is to develop an estimate of the prospective ILUV for the losses arising from the natural resource injuries described in Chapter 2. In situations where decisions must be made regarding design features or choices of statistical assumptions, we have adopted, within the economic framework necessary for measuring aggregate ILUV, the NOAA Panel recommendation as a desired philosophy for making these types of judgments: "Generally, when aspects of the survey design and the analysis of the responses are ambiguous, the option that tends to underestimate willingness to pay is preferred" (Arrow *et al.*, 1993; p. 4612).

The summary statistic we have chosen as an estimate of prospective ILUV is based on the Turnbull (1976) non-parametric, maximum likelihood (ML) estimator for interval-censored data. The Turnbull estimator uses respondents' choices to construct an interval estimate for the latent willingness to pay implied by each respondent's choice. As noted above, an individual's answer to a single question will distinguish either a lower or an upper bound for his or her WTP. By combining respondents' choices, we obtain estimates for the relative frequency of responses at different WTP intervals, $(0, W1AMT_i)$ and $(W1AMT_i, \infty)$, where $W1AMT_i$ is one of the five W-1 tax amounts administered to the different sub-samples. The first pair, $(0, W1AMT_i)$, defines the interval identified by $W1AMT_i$ as an upper bound and, the second pair, $(W1AMT_i, \infty)$, with $W1AMT_i$ as a lower bound. The six intervals or "steps" defined by W1AMT are: (1) \$0 to \$10, (2) \$10 to \$25, (3) \$25 to \$80, (4) \$80 to \$140, (5) \$140 to \$215, and (6) above \$215.

Two summary statistics, related to the sample mean, can be defined based on the Turnbull estimates of the fraction of the sample in each of the six intervals. The first of these we will refer to as the lower-bound mean. It is calculated by first assuming that the fraction of the sample estimated to be in each interval has a willingness to pay value equal to the lower end-point of the

interval and then estimating the ordinary sample mean. The second of these summary statistics is

the upper-bound mean. It is calculated in a similar manner by placing the fraction of respondents

estimated to be in an interval at the high end-point of the interval and then calculating the ordinary

mean. The unobserved sample mean is always bounded below by the lower-bound mean and

above by the upper-bound mean if there are identical subsamples at each of the tax amounts

asked.

It is important to recognize that any estimate of the sample mean which is lower than the

Turnbull lower-bound mean estimate or higher than the Turnbull upper-bound mean estimate is

inconsistent with the observed choices made by respondents.   Without additional statistical

assumptions, any observed choice measure is uninformative about where, within the two Turnbull

---

For instance, if 20% of the sample is estimated to be in the interval $10 to $25, the lower-bound
mean is calculated by assuming that this 20% of the sample is willing to pay exactly $10.

The upper-bound mean is potentially infinite unless reasonable additional assumptions (such as
no respondent would be willing to pay more than some fraction of his or her income) are imposed.

This statement is true irrespective of the particular tax amounts used to define the intervals,
although the particular tax amounts used can influence how much *less* the lower-bound mean is
than the sample mean and how much *greater* the upper-bound mean is than the sample mean.
Random assignment of respondents to tax amounts will result in subsamples at each tax amount
which are approximately equivalent in finite samples. The standard error of the lower-bound
estimate reflects possible variation in this estimate due to sampling variability.

In this regard, it can be seen that the Turnbull estimate of the distribution encompasses
parametric estimates of the sample mean which are consistent with the observed choices.

It is common practice in the literature to assume a specific parametric functional form to
describe the shape of the WTP distribution. Assuming a particular distributional specification
such as the log-normal is usually equivalent to assuming how the fraction of respondents
estimated to be in each Turnbull interval are arrayed within that interval. An estimate of the
sample mean derived using a parametric functional form will be sensitive to the specific parametric
distributional form assumed (and particularly the shape of the right tail associated with that
parametric distribution). One strategy for reducing this sensitivity to the assumed parametric
distribution has been to use the median as a measure of the central tendency for the estimated
distribution as this measure. However, from the perspective of the economic theory underlying

bounds, the sample mean lies. The most conservative assumption which is consistent with the observed choice measure is that the sample mean is equal to the Turnbull lower-bound mean.

§ 9.4 Univariate (Turnbull) Estimation of Lower-Bound Mean WTP

Table 9.4 reports the Turnbull estimate for the WTP distribution using the **W1CH** choice measure. Note that the third column in Table 9.4 (labeled "Probability of Voting *For* at Upper-Bound") is simply the estimated fraction of those in Table 9.3 who would vote *for* the program at each W1AMT. The elements in the table describe the intervals defined by W1AMT and respondents' choices. For example, we know a respondent's willingness to pay for the accelerated recovery program is greater than or equal to $10 if the respondent voted *for* the program at $10. If, on the other hand, a respondent voted *against* the program at $10, we know that the respondent's willingness to pay is less than $10 and possibly $0. Likewise, for a respondent who was asked about $80, a vote *against* the program implies that the respondent's willingness to pay for the accelerated recovery program lies somewhere in an interval from $0 to $80, while a vote *for* implies a maximum willingness to pay of at least $80. In this way, we can classify each respondent's willingness to pay into an interval depending on the W1AMT the respondent received.

---

the measurement of aggregate ILUV discussed in Chapter 3, the mean is clearly the preferred measure.

  The lower-bound mean recognizes that the lowest point in an interval is the threshold trade-off isolated by respondent choices.

  We assume that no respondent would demand compensation for implementing the accelerated recovery plan; that is, that no respondent has a negative WTP.

Table 9.4 Turnbull Estimate of WTP Distribution and Lower-Bound Mean:
W1CH Choice Measure [N=1857]

Lower Bound of Interval	Upper Bound of Interval	Probability of Voting *For* at Upper Bound	Change in Density
$0	$10	0.559	0.441
$10	$25	0.463	0.096
$25	$80	0.329	0.134
$80	$140	0.265	0.064
$140	$215	0.223	0.042
$215	∞	0.000	0.223

Log-Likelihood -1155.65
Estimate of lower-bound mean $63.24
Standard error of the estimate $2.54

It is important to recognize that the Turnbull estimator does not assume that a respondent who votes *against* at $80 is willing to pay $0; rather, the Turnbull estimate for the intervals of the WTP distribution identified by W1AMT can be thought of as being sequentially built up. The fraction of respondents voting *for* at $10 identifies the probability of voting *for* the accelerated recovery program at the upper bound (0.559) and the first entry in the change in density column is the percent of respondents voting *not for* (0.441=1.000 - 0.559). The fraction of respondents (0.463) voting *for* at the second tax amount, $25, defines the percent willing to vote *for* at least $25 and the percent (0.096) willing to vote *for* amounts between $25 and $10. This latter percent is calculated by taking the difference between the percent willing to vote *for* $10 (*i.e,* 0.559) and the percent willing to vote *for* $25 (*i.e.,* 0.463). This difference (0.096) appears in the second row of the change in density column. Likewise, the percent willing to vote *for* at $80 defines the percent willing to vote *for* at $80 shown in the third column (0.329) and the difference (0.134)

between the percent willing to vote *for* at $25 and $80 defines the percent willing to vote *for* an amount between $25 and $80 and is displayed in the fourth column. The rest of the table is built up in an analogous fashion.

The lower-bound estimate of the mean, $63.24, is obtained by assuming that all of the fraction of the sample estimated to be in a particular interval falls at the lower end of that interval. For example, respondents who voted *against* at $10 and thus fall into the [$0—$10] interval are assumed to have a willingness to pay of $0. Respondents who voted *for* at $215 and thus fall into the [$215—∞] interval are assumed to have a maximum willingness to pay of $215. The standard error of the estimated mean is fairly small, $2.54, indicating reasonable precision in this estimate.

§ 9.5 Bivariate Relationships Including NOAA Panel Cross-Tabulations

The NOAA Panel recommends that summaries of the responses to the "primary valuation" question (for our survey, embodied in the **W1** and **W1CH** choice measures) be broken down into categories to facilitate interpretation. The recommended categorical items include income, prior knowledge of the site, variables related to prior interest in the site such as visitation rates, distance

---

The numbers in the change in density column are the actual parameter estimates from the Turnbull model. Note that because the sum of the changes in density estimates must equal one, the last change in density (0.223) is not directly estimated but rather is calculated as one minus the sum of the first five changes in density. The z-statistics for the five change in density parameters estimated by the Turnbull procedure are 17.18, 2.59, 3.71, 1.92, and 1.35. The calculation of the standard error for the lower-bound mean estimate is described in Appendix E.

The lower-bound estimate of the mean is calculated by multiplying the lower bound of the interval column by the change in density column and then summing these amounts. In this instance, [$0 x 0.441] + [$10 x 0.096] + [$25 x 0.134] + [$80 x 0.064] + [$140 x 0.042] + [$215 x 0.223] = $63.24.

The corresponding estimate for the double-bounded choice measure taking account of changes, **WDBCH**, yields a somewhat higher lower-bound estimate of the mean of $67.69 (s.e. $2.92) and is displayed in Table F.7 in Appendix F.

to the site, attitudes toward the environment, attitudes toward big business, understanding of the task, belief in the scenario, and ability/willingness to perform the task.

This recommendation is addressed in three ways. In this section, we report the cross-tabulation results for responses to questions designed to address each of the recommended items with the choice measures, **W1** and **W1CH**. These cross tabulations test whether the *for* and *not for* choices are influenced by each of the variable's responses. Second, for illustrative purposes, we consider one at a time how a subset of these factors would influence the Turnbull estimate of the lower-bound mean. Third, we present a multivariate analysis in the following section which addresses both a subset of the items recommended by the Panel and others hypothesized to influence respondents' choices.

Table 9.5 describes the specific source of the information used in each of the cross-tabulations. In most cases, these correspond to questions in the main study survey. In a few cases, the measure was constructed using two or more variables measured by the survey. The table also includes a short descriptive summary of the information and an indication of whether the source of the information directly (D) or indirectly (I) measures the item identified by the Panel. As shown in the table, the survey instrument contains multiple variables for some of the Panel's recommended items.

Table 9.5 Description of Sources of Information for Cross Tabulations

Recommended Information	Source	Category[a]	Description
Income	C-13	D	Total household income before taxes in 1993
Prior Knowledge of Site	A-11	D	Heard about the DDT and PCB deposit off South Coast
Prior Interest in the Site	C-2 C-3	D	Have visited Catalina or other Channel Islands; Most recent visit in last five years

Recommended Information	Source	Category[a]	Description
	Saltwater Recreation[b]	I	Saltwater boating, fishing, or going often to the beach
	B-13	I	Bird watcher
Attitudes Toward the Environment	A-1b A-1e A-2e	D	Reducing air pollution in cities; Protecting coastal areas from oil spills; Protecting endangered wildlife species
	B-17	I	Respondent's self-evaluation on environmentalist scale
Attitudes Toward Big Business	C-24f	D	Confidence in large corporations
	C-24c	I	Confidence in scientists who work for industry
Distance to the Site	SOUTH COAST PSU's	D	Location of respondent's residence in PSU's comprising affected South Coast area
	FARNORTH PSU's	D	Location of respondent's residence in PSU's north of San Francisco Bay area (farthest area from that affected)
Understanding of the Task	D-1c	I	Interviewer evaluation of respondent's understanding of material presented in A-3 through A-16
	D-2	I	Interviewer evaluation of respondent comments indicating difficulty in understanding natural recovery or accelerated recovery program
	D-3	I	Interviewer evaluation of respondent understanding of voting questions
Belief in the Scenario	B-1	D	Respondent agrees DDT and PCBs could cause reproductive problems
	Natural recovery: more time[c]	D	Respondent judgment about timing of natural recovery
	Natural recovery: less time[c]	D	Respondent judgment about timing of natural recovery
	Accelerated Recovery Program: works[d]	D	Respondent judgment about effectiveness of accelerated recovery program
	Accelerated Recovery Program: not works[d]	D	Respondent judgment about effectiveness of accelerated recovery program
	B-6	D	Respondent judgment about limit of special tax to single year

216

Recommended Information	Source	Category[a]	Description
Ability/Willingness to Perform Task	D-4	I	Interviewer evaluation of whether respondent impatient to complete interview

"D" indicates a *direct* connection between the question and the information sought in the NOAA Panel's recommendati
"I" designates an *indirect* connection.
SWATREC indicates participation in saltwater recreation in the form of fishing or boating in the last five years (B-9 an
11, respectively) or often going to the beach (B-15).
Questions B-2 and B-3 are used to construct (0,1) indicator variables for whether respondent felt natural recovery woul
MORETIME or LESSTIME.
The PWORKS variable (describing whether the program works) is formed by combining those answering "yes" to
"completely effective in solving the reproduction problem in five years" in B-4 with those answering "mostly effective"
B-5 follow-up question. These responses were coded as 1 and 0 otherwise. The PNOTWORK variable describing an
ineffective program was formed by setting B-5 responses corresponding to "not too effective" or "not effective at all" eq
1 and 0 otherwise.

Table 9.6 summarizes the cross-tabulation results, including the p-values, for the **W1** and

W1CH choice measures. For the cross-tabulations shown in the table, the null hypothesis is

whether the distribution of responses *for* and *not for* the accelerated recovery program is affected

by the categories used to describe each of the matched source variables. For example, in the case

of income, the null hypothesis concerns whether respondents' choices are affected by the reported

income category. The reported p-value is the probability that the test result would call for

incorrectly rejecting a "true" null hypothesis of no association between the choice measure and the

source variable. The last column in the table reports the decision — assuming a p-value of 0.05

— that would be made about differences in the distribution of responses between *for* and *not for*

choices (using both the **W1** and **W1CH** choice measures) and the categories in each of the

---

See Appendix D-3 for a presentation of each of the cross-tabulation tables.

Three p-values are commonly used for deciding whether to reject/not reject statistical
hypotheses: 0.10, 0.05, and 0.01. Of these, 0.05 is perhaps the most commonly used criteria.
The p-value of 0.10 tends to be used in smaller samples where there is less statistical power to test
a hypothesis or when there is a lower risk involved in rejecting the null hypothesis when it is true.
A p-value of 0.01 is sometimes used when there is a higher risk involved in falsely rejecting the
null hypothesis. Using the statistic's actual p-value, it is possible to evaluate how sensitive this

information variables. The label "R" indicates that the null hypothesis was rejected (thus implying some association between choices and the information variable), and "N" indicates that the null was not rejected (hence suggesting no association).

Table 9.6 Cross-Tabulation Summary

Recommended Information	Source[a]	Choice Measure	p-value[b]	Reject/Not Reject Hypothesis of No Association
Income	C-13	**W1**	0.00	R
		W1CH	0.01	R
Prior Knowledge of Site	A-11	**W1**	0.04	R
		W1CH	0.03	R
Prior Interest in the Site	C-2	**W1**	0.28	N
		W1CH	0.90	N
	C-3	**W1**	0.25	N
		W1CH	0.07	N
	Saltwater Recreation	**W1**	0.00	R
		W1CH	0.00	R
	B-13	**W1**	0.00	R
		W1CH	0.00	R
Attitudes Toward Environment	A-1b	**W1**	0.00	R
		W1CH	0.00	R
	A-1e	**W1**	0.00	R
		W1CH	0.00	R
	A-2e	**W1**	0.00	R
		W1CH	0.00	R
	B-17	**W1**	0.00	R
		W1CH	0.00	R
Attitudes Toward Big Business	C-24c	**W1**	0.66	N
		W1CH	0.60	N
	C-24f	**W1**	0.46	N
		W1CH	0.34	N
Distance to Site	SCOAST	**W1**	0.02	R
		W1CH	0.03	R
	FARNORTH	**W1**	0.01	R

decision is to the selection of a particular p-value.

Recommended Information	Source[a]	Choice Measure	p-value[b]	Reject/Not Reject Hypothesis of No Association
		W1CH	0.01	R
Understanding of Task	D-1c	**W1**	0.13	N
		W1CH	0.09	N
	D-2	**W1**	0.68	N
		W1CH	0.24	N
	D-3	**W1**	0.17	N
		W1CH	0.68	N
Belief in Scenario	B-1	**W1**	0.02	R
		W1CH	0.01	R
	Natural Recovery: More Time	**W1**	0.00	R
		W1CH	0.00	R
	Natural Recovery: Less Time	**W1**	0.00	R
		W1CH	0.00	R
	Accelerated Recovery Program: Works	**W1**	0.00	R
		W1CH	0.00	R
	Accelerated Recovery Program: Not Effective	**W1**	0.00	R
		W1CH	0.00	R
	B-6	**W1**	0.00	R
		W1CH	0.00	R
Ability/Willingness to Perform Task	D-4	**W1**	0.00	R
		W1CH	0.00	R

The source is the question number in the main survey unless otherwise indicated; see preceding table. Refused/
sure/not ascertained categories have been set to missing for the source variables and excluded from the cross
tabulations.

The p-value is the probability level estimated for a Type-I error for a χ^2 statistic using a cross-tabulation of the
measure and the recommended information variable.

These cross-tabulations permit a simple test of association between respondents' choices

and three different types of information. The first type relates to the characteristics and attitudes

of respondents. Here we would expect to see differences in the choice measures with respect to

at least some respondent characteristics. These results suggest that in each group of variables, except *attitudes toward big business* and *understanding of the task*, at least one measure in the group is significantly related to the choice regarding the accelerated recovery program. These include variables with direct economic interpretations such as income (C-13), as well as measures of activities that might be hypothesized to be related to the injured resources, such as participation in various forms of saltwater recreation, and identifying bird species (B-13). The choice measure used in the cross-tabulation, **W1** or **W1CH**, does not influence this conclusion.

Environmental attitudes are consistently related to differences in the decisions about the program. These include those variables from survey questions asked before the program and injuries are described (A-1b, A-1e, and A-2e) and a later question which asks for a general self-evaluation on an environmentalist scale (B-17). Attitudes toward big business are represented in two ways: first, in the degree of confidence in scientists who work for industry (C-24c), and second, in the degree of confidence in large corporations (C-24f). Neither is significantly related to choosing the accelerated recovery program.

Prior knowledge does appear to be related to respondents' choices. *Distance to the site* as measured here by SCOAST, a dummy variable for respondents whose residences fall within the PSU's in the South Coast area (*i.e.,* Los Angeles and Orange counties), was also found to be associated with choices. There was a difference in the opposite direction for FARNORTH, PSU's north of the San Francisco Bay Area, the area farthest away from the South Coast.

The other two types of information are respondent's *understanding of the task* and his/her *belief in the scenario*. Interviewers' evaluations of respondents' performance (used as an indirect measure for respondents' *understanding of the task)* were not a factor in distinguishing the pattern of choices. However, as we will see below, a broader definition of possible interviewer-identified

problems is associated with a smaller estimate of the Turnbull lower-bound mean. As we would

expect, all of the measures of *belief in the scenario*, measured by both respondents' perceptions of

the natural recovery process and the effectiveness of the accelerated recovery program,

distinguish the pattern of choices. The acceptance of the one year limit to the special tax is also a

distinguishing feature. The interviewers' evaluations of respondents' impatience to complete the

interview offer an indirect gauge of their willingness to perform the task. In this case, the cross-

tabulation suggests a significant association with the impatient respondents tending to vote *not for*

the program.

It is possible to use the variables identified by the NOAA Panel as a basis for dividing the

base sample into sub-samples. Separate Turnbull estimates can be computed for each sub-

sample's WTP distribution, and the lower-bound means compared. As a rule, repeating this

process for the categorical variables defined by the survey and discussed earlier (in terms of cross-

tabulations) indicates significant differences in the estimated lower-bound means across these

categories. For example, splitting the sample according to whether respondents were interviewed

in the South Coast area yields, as we would expect, a significantly greater estimate of the **W1CH**

lower-bound mean for SCOAST households in comparison with FARNORTH households (*i.e.,*

$71.65 versus $43.26; t=3.25, p < 0.001). Households with bird watchers have a lower-bound

estimate for mean WTP of $76.12 versus $53.97 for those who do not (t=4.34; p < 0.001).

In several instances, the interesting way to split the data into two sub-samples (*e.g.*, the 19
respondents who interviewers identified as having understood the injury and program material
only slightly or not at all versus the rest of the respondents) results in at least one very small sub-
sample and hence fairly unreliable estimates. As a consequence, the computations presented
should be taken as illustrative of the implications of the differences tested with the cross
tabulations.

Extreme splits such as comparing those on question A-2e who think spending on endangered
wildlife programs should be increased a great deal to those who think spending should be reduced

When comparable sample splits were considered for other elements in Table 9.6, the relationships between the lower-bound means estimated from each distribution were consistent with our prior expectations. For example, among those respondents who expressed a belief that natural recovery would take more time, we would expect that the estimated mean derived from a WTP distribution based on their choices would be significantly larger than that estimated from those who indicated otherwise; this is indeed the case ($99.69 versus $60.88; t=3.73; p < 0.001). These relationships reinforce the test results derived from comparing the cross-tabulations recommended by the NOAA Panel.

It is also possible to look at splitting the sample into those respondents with possible interviewer-identified problems according to the PINTPROB measure defined in section 8.5, and those not identified as having problems. We find that respondents with possible interviewer-identified problems have a lower-bound estimate of the mean which is approximately one half the size of the corresponding estimate for respondents identified as not having problems ($32.88 versus $64.48; t=-3.08; p=0.002).

---

a great deal produce quite large differences ($111.42 versus $18.20; t=7.19; p < 0.001).

§ 9.6 Construct Validity Using a Multivariate Approach

The estimation of a multivariate choice function is a statistical method used to relate respondents' choices to their evaluations of the accelerated recovery program as well as to their demographic characteristics and attitudes. These functions are often used to demonstrate *construct validity*, one of the standard validity concepts widely accepted for use in evaluating models. Construct validity refers to the degree to which a measure relates to other measures predicted by theory. In examining construct validity, we look at whether variation in the **W1CH** choice measure is systematically related to factors suggested by economic theory such as preferences for the object of choice, the cost of program, and the ability to pay for it. Other factors relevant for this application include measures of respondents' evaluations of the injuries and the characteristics of the accelerated recovery program. For example, we should expect those respondents who thought natural recovery would take less time than was described to them in the questionnaire to be less likely to vote *for* the program.

When we move from this general description of an evaluation of construct validity, it is important to acknowledge that most predictions are general and simply indicate whether the direction of the association between a variable and a respondent's choices should be positive or negative. Equally important, they rely on observing a large enough number of individuals with the characteristics or attitudes hypothesized to be related to a choice to be able to estimate the parameter of interest with reasonable precision.

Mitchell and Carson (1989) discuss two forms of construct validity: convergent validity and theoretical validity. The former refers to whether the measure of interest is correlated with other measures of the same theoretical construct and is not directly applicable here.

Because respondents' attitudes and demographic characteristics are measured in several different ways, and their evaluations of the injuries and the program are also evaluated from different perspectives, it is reasonable to expect correlation among these sets of variables. This

§ 9.6.1 Definition of Covariates in Choice Function

Table 9.7 presents a multivariate choice function estimated using a probit model with **W1CH** as the dependent indicator variable. We used a probit model because of its simplicity in estimation and presentation, and because it is one of the models most frequently used to relate a binary, discrete-choice variable to a set of possible predictor variables.

**Table 9.7 Multivariate Analysis of Construct Validity:
Probit Estimates for W1CH Choice Valuation Function**

---

multicollinearity can reduce the ability of the model to distinguish individual effects precisely when there are multiple measures reflecting closely-related influences on respondents' choices.

Missing values for income have been replaced with an estimate based on the median income in the 1990 Census block, housing type, education, gender, race, age, and qualitative variables for the number of employed adults in the household. Appendix F reports the model for estimating income (Tables F.1 and F.2), more detailed definitions of the variables included in the choice function (Table F.3), as well as the model presented in Table 9.7 but excluding the households who do not report income from the sample (Table F.4). Doing this does not change the sign or significance of the income measures or the role of any other variables. It does reduce the sample from 1857 to 1692 so the p-values for some of the tests for relationships between these variables and respondents' choices necessarily decrease. The most notable examples arise with the location variables.

Generalizing the Turnbull estimate with covariates requires adding parametric structure either in the form of the distribution assumed to give rise to a probability distribution for the choice measure or the index function used to describe how the covariates influence choices, or both. Because the multivariate analysis was intended to test construct validity, we used a simple format for these tests and evaluated the sensitivity of the results to adjustments for outlying observations and flexibility on the parametric restrictions imposed on W1AMT and income, the two variables that were not qualitative (*i.e.*, categorical) variables. We considered a version of the probit estimator due to Pregibon (1982) to adjust for outlying observations. The resulting model is similar to that reported in Table 9.7. We also considered the generalized additive form of the probit model due to Hastie and Tibshirani (1990). This version uses a smoothing spline technique for the two continuous variables, W1AMT (which allows more flexibility in the error term) and income. The generalized additive form did not significantly improve the fit over that of the probit reported in Table 9.7. Neither approach altered our conclusion on construct validity.

Variable	Coding	Parameter Estimate	Z-Statistic	p-value (two-sided)	Variable Mean
CONSTANT	Equals 1 of all respondents	-1.1592	-1.86	0.063	—
LW1AMT	Log of W1AMT	-0.4025	-12.77	0.000	4.0616
LINC1	Log of income if < median California household income ($35,173); 0 otherwise	0.1745	2.86	0.004	5.0777
LINC2	Log of income if ≥ to $35,173 and < $150,000; 0 otherwise	0.1491	2.70	0.007	4.85489
LINC3	Log of income if ≥ $150,000; 0 otherwise	0.1142	2.17	0.030	0.4493
EDUC	College Associates degree or higher=1; 0 otherwise	-0.1770	-2.26	0.024	0.3802
NOTAX	Did not pay California taxes=1; 0 otherwise	0.4854	3.43	0.000	0.1077
COASTIP	A-1c protect coastal area extremely important=1; 0 otherwise	0.1486	2.00	0.046	0.3667
COASTNIP	A-1e protect coastal area not important=1; 0 otherwise	-0.7135	-1.38	0.169	0.0135
WILDSP	A-2e increase endangered wildlife spending=1; 0 otherwise	0.4180	5.30	0.000	0.4847
WILDNSP	A-2e decrease endangered wildlife spending=1; 0 otherwise	-0.2676	-2.15	0.032	0.1486
NONSENV	B-17 not at least a somewhat strong environmentalist=1; 0 otherwise	-0.2404	-3.09	0.002	0.3861
MORETIME	B-3 natural recovery a lot more time=1; 0 otherwise	0.5250	3.63	0.000	0.0630
LESSTIME	B-3 natural recovery a lot less time=1; 0 otherwise	-0.2915	-2.69	0.007	0.1497
PWORKS	B-4 and B-5 expect program to be completely or mostly effective=1; 0 otherwise	0.5998	7.46	0.000	0.6317
PNOTWORK	B-5 expect program to be not too effective or not at all effective=1; 0 otherwise	-1.2578	-4.77	0.000	0.0872

Variable	Coding	Parameter Estimate	Z-Statistic	p-value (two-sided)	Variable Mean
QUESPROG	At A-13 or A-15 asked question about how program worked or its cost explanation=1; 0 otherwise	-0.2961	-3.58	0.000	0.2439
PAYMORE	B-6 does not think will only have to pay special tax for one year=1; 0 otherwise	-0.2817	-3.67	0.000	0.3667
PAYVEH	C-26 prefer tax vehicle over higher prices=1; 0 otherwise	0.3908	5.44	0.000	0.3533
CONFCGV	C-24e great deal of confidence in California State Government=1; 0 otherwise	0.3119	1.72	0.085	0.0393
NCONFCGV	C-24e no confidence in California government=1; 0 otherwise	-0.2065	-2.13	0.033	0.1745
LOWSPEND	Wants increased spending only on one or no programs (A-2a, A-2b, A-2c, A-2d and A-2f)=1; 0 otherwise	-0.3230	-3.72	0.000	0.2606
SWATREC	B-10, B-12, B-15 participate in saltwater boating or fishing or often go to beach=1; 0 otherwise	0.2160	2.91	0.004	0.5859
BIRDWATC	B-13 birdwatcher=1; 0 otherwise	0.1790	2.41	0.016	0.4136
TVBIRDS	B-14 often watch tv programs about animals and birds=1; 0 otherwise	0.1861	2.52	0.012	0.4416
EATFISH	B-16 household often eats fish=1; 0 otherwise	0.1759	2.45	0.014	0.4146
SCOAST	Los Angeles or Orange Counties=1; 0 otherwise	0.1668	2.17	0.030	0.3253
FARNORTH	North of San Francisco Bay Area=1; 0 otherwise	-0.2468	-2.10	0.036	0.1147

N = 1857
Log (L) = -879.78
Pseudo R^2 = 0.279

In considering the model presented in Table 9.7, first note that the randomly assigned treatment variable, W1AMT, entered as a natural log, is a negative determinant of respondents' decisions about the accelerated recovery program and is a highly significant determinant of respondents' choices in the model. The other variables selected for inclusion in the choice model can be grouped into five broad categories following the general format of the NOAA Panel's recommendations for cross tabulations: variables measuring either directly or indirectly respondent economic characteristics; respondent preferences and demographic characteristics; respondent evaluations of the injuries and accelerated recovery program; respondent interpretations of the choice mechanism and circumstances of the choice; and respondent interest in, use of, and proximity to the affected natural resources. We now turn to a specific discussion of the other variables in the construct validity equation.

The first group of variables, which measure economic characteristics, include income, EDUC, and NOTAX. As noted in the table, three income classes are identified, those below the median, annual California household income of $35,173 (INC1), between $35,173 and less than $150,000 (INC2), and those of $150,000 and above (INC3). The next variable, EDUC, is a qualitative variable indicating that the respondent has at least an Associates degree from an academic college program. NOTAX is an indicator that the household did not pay any California income taxes last year.

A series of five environment variables follow. The first four are defined from the initial questions asking respondents about preventing oil spills in coastal areas (A-1e) and protecting endangered wildlife species (A-2e). The former variables (COASTIP and COASTNIP) are entered as qualitative variables identifying those respondents who at A-1e rated the issue as

"extremely important" as well as those indicating it was "not important at all." Each category is measured relative to the middle group. In the case of endangered wildlife (A-2e), the two variables are defined in terms of the size of spending changes, with those desiring spending increases (WILDSP) and spending decreases (WILDNSP) identifying the separate qualitative variables. The fifth environmental variable, NONSENV, identifies individuals who would *not* consider themselves to be at least a somewhat strong environmentalist.

The next set of factors are related to the program. The first two variables, MORETIME and LESSTIME, relate to respondents' evaluations of the length of natural recovery. The next two variables, PWORKS and PNOTWORK, relate to respondents' evaluations of whether the program would be effective. QUESPROG is an indicator of whether the respondent asked a question(s) about how much the program would cost or how it would work in A-13 or A-15. The last variable, PAYMORE, relates to whether the respondent thought the tax payment might not be limited to one year.

The next group of variables relates more generally to respondents' views about government programs. The first, PAYVEH, is defined from respondents' evaluations of whether taxes were the appropriate way to pay for new programs to protect the environment. The next two, CONFCGV and NCONFCGV, reflect levels of trust in the California state government. Here the coding of C-24e into two categorical (0,1) indicator variables, distinguishing respondents who express a great deal of confidence in California state government and those with no confidence from the rest of the respondents. The last variable in this group, LOWSPEND, is an indicator variable for respondents who express willingness to increase spending on none or, at most, one of the programs asked about in question A-2 (*i.e.*, new state prisons, public transportation in Los Angeles, raises for state college and university faculty, shelters for the

homeless, and more lifeguards at state beaches).

The last category of variables identifies the use activities that may be "related" to the injured natural resources, including: saltwater recreation (SWATREC, *e.g.*, saltwater fishing, boating, or beach use), often eating fish (EATFISH), bird watching (BIRDWATC), and watching television programs about birds and animals in the wild (TVBIRDS). Finally, we have also included in this group two qualitative variables identifying the respondent's location in relationship to the area closest to the natural resource injuries. Here FARNORTH and SCOAST are distinguished from the rest of the State.

§ 9.6.2 Interpretation of Covariates in Choice Function

In each case, the variables hypothesized to influence choices do so in the ways anticipated and the estimated coefficients are generally significant at p-values less than conventional standards. Distinguishing the extreme attitudes is helpful in highlighting these differences. The pattern of positive and negative signs on the coefficients suggests that those with attitudes supporting the environment, expressed as either a concern or a willingness to increase spending to protect endangered wildlife, are *more* likely to vote for the program. Those respondents who do not consider themselves strong environmentalists, who prefer not to increase spending on public programs, and who do not trust California government are *less* likely to support the program. EDUC has a negative effect on a respondent's willingness to support the program and is consistent with some skepticism toward the program expressed by educated respondents in focus groups.

Note that Table 9.7 reports p-values for two-sided hypothesis tests. In most instances, the hypothesis about the coefficient on a particular test is of the one-sided form (*e.g.,* a null hypothesis that respondents who do not think the program works are as likely to vote *for* the

As noted above, there are five program-related variables. We would expect those respondents who thought that natural recovery would take a lot longer than fifty years (MORETIME) to be more likely to vote *for* the program while those who thought that natural recovery would take a lot less than 50 years (LESSTIME) to be more likely to be *not for* the program. The coefficients in Table 9.7 confirm this; both effects are statistically significant. It is interesting to note that the two effects are almost exactly offsetting.

We would also expect that those who thought the program would work (PWORKS) would be more likely to vote *for* while those who thought the program would not work (PNOTWORK) would be less likely to vote *for* the program. Again this is the case and the effects are highly significant. Since ideally all respondents would be in the PWORK=1 category, the overall effect here is to lower the percentage of *for* votes. The coefficient on QUESPROG, a related variable, is negative and significant. To the extent that the questions asked reflect skepticism about the program, rather than cost sensitivity, the effect is undesirable. We would also expect those who did not think they would have to pay the amount asked about for only one year (PAYMORE) to be less likely to vote *for* the program. This is the case, and the effect is to significantly reduce the likelihood of a *for* vote. Taken as a whole, deviations from believing that there would be a 50 year natural recovery, that the accelerated recovery program would work, and that the special tax would only have to be paid for one year results in a significantly lower probability ($p < 0.001$) of voting *for* the program.

---

program as other respondents versus the alternative that they are less likely). For one-sided hypothesis tests, the reported (two-sided) p-values should be divided by 2.

  The absolute value of the coefficient on MORETIME is almost twice that of LESSTIME. However, the percent of respondents giving a LESSTIME answer is more than double that of those giving a MORETIME answer.

The next group of variables, PAYVEH, CONFCGV, NCONFCGV, and LOWSPEND, all have the expected signs and are significant. Those favoring the use of government taxes to effect environmental improvements are *more* likely to vote for the program as are those indicating more trust in the government. Those with no confidence in the government and those not favoring increased government spending in general are *less* likely to vote for the program.

The next group of variables, SWATREC, BIRDWATC, and EATFISH, are related to the natural resource. They are all positive and significant. That is, individuals whose activities and interests are related to saltwater recreation (SWATREC) and wildlife (BIRDWATC and TVBIRDS) are *more* likely to vote for the program. The same is true for households who often eat fish. Respondents living in the FARNORTH PSU's are *less* willing to vote for the program and those in the SCOAST PSU's *more* willing to vote for the program than those in the rest of the state.

All of these factors are quite robust determinants of **W1CH**. We considered several different codings of these responses, such as including only one side of on attitude scale in comparison to all others (instead of using the two extreme evaluations relative to the intermediate ones). In most cases, the resolution for these variables was improved by accounting for the extremes in attitudes relative to intermediate opinions. However, the basic conclusions remain the same under reasonable alternative coding schemes. Moreover, they are not sensitive to the estimator used to describe the determinants of these initial responses. A Weibull choice model using **W1CH** rather than the log-normal probit model in Table 9.7 or their double-bounded counterparts assuming Weibull or log-normal distributions (see Table F.8 in Appendix F) also yield the same basic conclusions.

Household income was the variable most sensitive to its specification in the model. The

model reported here allows income to have a different coefficient depending upon the level of household income. The coefficients for all three income terms are positive and statistically significant with p-values less than 0.05 and the p-values are less than 0.01 for LINC1 and LINC2. Likelihood ratio tests of the hypothesis that all income coefficients are jointly zero reject at p=0.010, using the sample with the imputed income for missing income values, and at p=0.016 in the model dropping these observations.

If we do not allow the income coefficient to vary with level of income, then the effect of the log of income on the likelihood of favoring the program is positive but no longer statistically significant. This conclusion holds regardless of the treatment of missing values for income. Nonetheless, this specification would be rejected in favor of the one reported in Table 9.7 at p=0.005 (p=0.007 without the imputed income cases) using likelihood ratio tests. That specification is able to isolate a significant, positive effect of income because it includes variables that capture the negative evaluations of government spending programs, environmental projects, and the accelerated recovery program on the part of a disproportionate share of the sample's highest income households. Thus, in the absence of a measure capturing these attitudes, it is possible to misinterpret the effect of income on respondents' choices. Equally important, by allowing for different income coefficients for these responses with the level of income, the model further distinguishes this group of high income households from others in the sample.

The last variable to be noted in the model is NOTAX, the indicator for households who do not pay California state income taxes. The coefficient on this variable is positive and highly

---

Using income rather than log (income) yields a negative and insignificant coefficient. This specification, however, can also be rejected in favor of a specification with three income terms using likelihood ratio tests (p=0.020 using imputed income for missing income values and p=0.054 dropping these observations).

significant (p < 0.001), indicating that this group is more likely to favor the program. Because it

is possible that their choices reflect a recognition that they would not have to pay the stated tax

amount, there is reason to question whether they have accepted financial responsibility for their

choices. The next section discusses an adjustment in the choice measure used for developing the

final estimate of prospective ILUV that treats the potential incentives to households not paying

California taxes in a conservative fashion.

The survey questionnaire collected additional information on demographics, knowledge,

and attitude/behavioral information. Our evaluation of the construct validity model reported in

Table 9.7 also considered these variables. In discussing the results from the analyses of these

other variables, it is important to acknowledge that the objective of our multivariate analysis is to

evaluate whether measures of the primary economic, attitude, program-related, resource-related

and demographic factors influenced respondent choices. In addressing this question of what

should be included in the model, the sensitivity of our findings to alternative definitions for the

included variables, as well as to variables that might have been included, was a key consideration

for the analysis. The summary to this point has considered the implications of the format used to

represent the included variables. We now turn to the variables in the survey that were not

reported as arguments in the Table 9.7 model.

The respondent's evaluation of the "seriousness of the injury" (B-8) is not included in the

final model. When it was included in the specification, our conclusions about the effects of the

tax amount, income, program-related variables, and several important demographic variables were

not affected but some of the environmental variables were no longer significant. Because this

---

Adding two indicator variables for B-8 equal 1 ("not serious at all") or 2 ("not too serious") and
B-8 equal 4 ("very serious") or 5 ("extremely serious") in the model in Table 9.7 substantially
increases the model's predictive power. Both of these variables are highly significant (t = -6.00

evaluation was requested after a respondent's choice, it is reasonable to assume that it conveys some of the same information as the choice itself (*i.e.*, respondents rating the injuries as "very" or "extremely" serious are more likely to vote *for* the program.) This measure is used in Chapter 10 to evaluate whether independent samples perceived a difference between the base and scope versions of the injury descriptions.

A number of demographic variables (age, gender, race/ethnicity, citizenship, the number of children under 18 in the household, and having grandchildren), were not significant determinants of choices when income, attitude, and program evaluation variables were included in the model. Some of these, such as age (negative relationship) and children (positive), were statistically significant in bivariate relationships with **W1CH**. Other variables which were significant in bivariate relationships with **W1CH**, but not significant in the model in Table 9.7, include C-1 (positive), the number of years lived in California, and, as noted earlier in Table 9.5, A-11 (positive), having heard about the DDT/PCB deposit. A variable measuring respondents' desire to move from Los Angeles/Orange County (C-5) was a positive and significant determinant of choices when included in the model in Table 9.7. This question, however, was only asked of SCOAST area respondents and the collinearity between the two variables substantially increases the estimated standard error for the SCOAST variable. As a consequence, we retained only SCOAST which is the relevant variable for the full sample.

Overall, the construct validity model includes a large number of the factors describing respondents, their economic characteristics, attitudes, and evaluation of the injuries and

---

and t = 7.68 respectively). The inclusion of these two variables makes the income variables more significant and the environmental variables less significant.

  A-9, having heard about DDT, and A-10, having heard about PCB's, were not significant in either bivariate relationships with **W1CH** or in the multivariate model in Table 9.6.

accelerated recovery program. Moreover, our overall conclusion on construct validity withstood variations in the format of the included variables measuring the factors hypothesized to influence choices as well as to the inclusion of other potential determinants of these decisions.

§ 9.7 Correction for Non-Taxpayers

As noted in section 5.4, the payment vehicle used in this study is a one-time increase in California income taxes. Some respondents not currently paying state income taxes may not take a tax payment obligation as seriously as those who do pay taxes. While there are other differences in respondents' interpretations of the elements of the choices—as the multivariate choice model indicates—there is an important difference between these effects and the one associated with not paying California income taxes. By retaining the varied interpretations of the accelerated recovery program, we are understating the propensity of respondents to be willing to vote for the program (at all tax amounts). Thus, the absence of an adjustment is consistent with a conservative estimate of ILUV.

In contrast, respondents who do not pay California taxes appear more willing to vote *for* the program (at all tax amounts). The most conservative adjustment for this tendency is accomplished by re-coding the 80 respondents who did not pay California income taxes in 1993 and who voted *for* the program given the **W1CH** choice measure to *not for* votes. This effectively sets the lower-bound mean estimate for this group of respondents to zero. We refer to this choice measure as **W1CHNT**.

There is further support for the use of this modified choice measure, **W1CHNT,** in interviewer ratings of the respondent's attention to the choice questions. Respondents who did not pay California income taxes were more likely ($p < 0.001$) on the PINTROB measure

introduced in Chapter 8 to be rated as lower quality interviews, to be considered impatient (D-4

and D-4a), or to be only "slightly serious" or "not at all serious" (D-5) in responding to the choice

questions. While the number of respondents in each of these undesirable categories is fairly small,

these evaluations arise disproportionately from this grouping of respondents who do not pay

California income taxes. Thus, this pattern is consistent with the hypothesis underlying our

adjustment to the choice measure; that is, these respondents may not interpret the choice question

in the same way as those who do pay California income taxes.

Table 9.8 reports the distribution of **W1CHNT** by W1AMT, while Table 9.9 reports the

Turnbull lower-bound estimate for this choice measure. As is the case for the **W1** and **W1CH**

choice measures, a $\chi^2_{(4)}$ test (111.02) for the **W1CHNT** measure also rejects the hypothesis

(p < 0.001) that responses are not sensitive to W1AMT. As shown in Table 9.9, the estimated

lower-bound mean for the **W1CHNT** choice measure is $55.58 with a standard error of $2.43.

This estimate, smaller than that from the **W1CH** choice measure ($63.24, with a standard error of

$2.54), represents yet another conservative adjustment to the lower-bound estimate of mean

WTP.

Table 9.8 W1CHNT Choice Measure by W1AMT

W1AMT	For	Not For
$10	50.3%	49.7%
$25	39.8%	60.2%
$80	29.9%	70.1%
$140	23.6%	76.4%

The z-statistics for the five change in density parameters estimated by the model are 19.24, 2.86, 2.80, 1.93, and 1.70, respectively. The corresponding double-bounded estimate for **WDBCHNT** (Appendix F, Table F.10) is $59.53 (s.e. $2.78).

$215	18.6%	81.4%
$\chi^2_{(4)} = 111.02; \, p < 0.001$		

Table 9.9 Turnbull Estimate of WTP Distribution and Lower-Bound Mean:
W1CHNT Choice Measure [N=1857]

Lower Bound of Interval	Upper Bound of Interval	Probability of Voting *For* at Upper Bound	Change in Density
$0	$10	0.503	0.497
$10	$25	0.398	0.105
$25	$80	0.299	0.099
$80	$140	0.237	0.062
$140	$215	0.187	0.050
$215	∞	0.000	0.187

Log-Likelihood -1112.10
Estimate of lower-bound mean $55.58
Standard error of the estimate $2.43

§ 9.8 Summary

Chapter 3 described the relationship between an individual's choice and the trade-off used to construct a monetary measure of the economic value implied by that choice. We argued that, with minimal assumptions, these choices could be used to isolate a lower-bound measure of each individual's willingness to pay for an object of choice. Our CV survey was designed to present a choice consistent with measuring prospective ILUV for the injuries described in Chapter 2. This was accomplished using an accelerated recovery program as the object of choice.

The statistical method used, the non-parametric maximum likelihood Turnbull estimator, allows the estimation of a lower-bound mean which is equal to, or lower than, the sample mean. The assumptions this estimator uses are simply those which underlie the choice model developed

in Chapter 3. That is, the Turnbull estimator does not require either distributional or preference-related assumptions, aside from the fundamental premise that respondents' choices are based on what is their most preferred alternative of the options presented to them.

To help gauge the reliability of respondents' choices in the base sample, two types of evaluations were presented. The first considered bivariate relationships, principally in the form of cross-tabulations between the information variables recommended by the NOAA Panel and the **W1** and **W1CH** choice measures. Overall, the bivariate analysis provided support for the presence of relationships that are hypothesized to be associated with choices for the accelerated recovery program.

The second evaluation considered the same issues but in a more structured format using a conventional, multivariate choice model. A probit model was estimated and used to help identify the determinants of the **W1CH** choices. These determinants include respondent economic characteristics, environmental variables, variables specifically related to the respondent's perception of the accelerated recovery program, general attitudes toward government programs, and variables which help identify the respondent's relationship to the injured natural resource. In each instance, factors hypothesized to be associated with the choices were found to be consistent with prior expectations and the relationships were statistically significant determinants of **W1CH**. Moreover, these are robust effects that do not seem to change much with the specific coding of the variables involved.

Only in the case of household income were the results found to be fairly sensitive to the assumed parametric specification. A specification that allows the probability of voting *for* the program to vary with three different income groups isolates a positive and significant effect between income and **W1CH**. Several simpler specifications which resulted in insignificant income

coefficients were rejected using likelihood ratio tests.

The lower-bound mean for respondents' willingness to pay for the accelerated recovery program was estimated to be $63.24. When the choice measure was adjusted to treat all households not paying California income taxes as having *not for* votes for the program, the lower-bound mean was $55.58. In both cases, the maximum likelihood estimates of the standard error for these lower-bound means were small (2.54 and 2.43, respectively) suggesting a reasonably high level of precision in these estimates.

§ 10 Responsiveness to Scope of the Injury

§ 10.1 Introduction

This chapter presents the empirical results of the comparison between responses to the base and scope survey instruments described in Chapter 6 and contained in Appendices A.1 and A.2. As noted in Chapters 5 and 7, these survey instruments were developed and administered to permit complete comparability between the results from the two independent samples.

Scope refers to how the willingness to pay constructed from respondents' choices responds to changes in the object of choice. The NOAA Panel suggested that economic measures of value should display differences for objects of choice that are recognized by respondents to be different. This view implies that respondents' WTP for different objects of choice should be different and it indicates a direction for this difference. That is, an object of choice understood by respondents to be larger should have an estimated WTP which is greater than the WTP for a desirable object of choice recognized to be smaller. A test of whether respondents are willing to pay more for a larger set of injuries than a smaller set is one of the principal tests of reliability recommended by the NOAA Panel.

While the NOAA Panel report does not provide a detailed description on how to evaluate this property, two general aspects of their recommended evaluation are important to the design of our scope test. First, the scope test should be administered to independent samples. This requirement follows from the Panel's discussion of issues associated with *embedding.* The Panel

---

The base version of the survey involves 2 species of fish and 2 species of birds with a 50 year natural recovery period. The scope version involves 2 species of fish with a 15 year natural recovery period. See Chapter 6 for a detailed description of the base and scope survey instruments.

As noted in Chapter 3, embedding has been used to refer to a number of different phenomena; the context here is that where one object of choice nests or encompasses another object of choice.

rejects the possibility of asking each respondent about several different objects of choice noting that:

> We must reject one possible approach [for dealing with embedding], that of asking each respondent to express willingness to pay to avert incidents of varying sizes; the danger is that embedding will be forcibly avoided, still without realism. [Arrow *et al.*, 1993; p. 4608; bracketed phrase added]

> Second, in describing the scope property, the NOAA Panel notes that respondents must be able to *recognize* the differences in the objects of choice presented: Rationality in its weakest form requires certain kinds of consistency among choices made by individuals. ... Common notions of rationality impose other requirements which are relevant in different contexts. Usually, though not always, it is reasonable to suppose that more of something regarded as good is better so long as an individual is not satiated. This is in general translated into a willingness to pay somewhat more for more of a good, as judged by the individual. [Arrow, p. 4604]

Because these different objects of choice are presented to independent samples, a judgment about respondents' understanding of differences in the objects must be based on a qualitative analysis undertaken as part of the development of the survey instrument. Chapter 5 summarizes this development process. This chapter describes the results of our scope test which confirm that respondents perceived differences in the objects of choice described and were, overall, willing to pay more for the program offered in the base version.

Before presenting these test results in detail, section 10.2 briefly reviews the sampling design underlying our scope test. Then, section 10.3 explores whether key predictor variables— constructed from questions administered prior to the description of the injuries—and demographic variables have similar response distributions across the base and scope samples. This section also addresses whether respondents from the independent samples were sensitive to the differences in the base and scope injuries by examining whether variables (other than WTP) which should be sensitive to scope are indeed sensitive.

There are several different choice measures which could be used in a test of sensitivity to scope. The single-bounded choice measures described in Chapter 9 were:

(a)   **W1**, the respondent's original response to the W-1 question coded as *for* or *not for*,

(b)   **W1CH**, the respondent's final choice after opportunities for reconsideration, and

(c)   **W1CHNT**, **W1CH** adjusted by setting the responses of households not paying California taxes to *not for*.

This section reports scope tests using **W1**, **W1CH**, and **W1CHNT**. Our principal focus is on the **W1CH** case, as this case represents the respondent's final choice (whereas the **W1CHNT** choice measure represents an artificial adjustment made to the data to reduce any chance of over-estimating WTP). Scope tests based on the double-bounded choice measures, **WDB**, **WDBCH**, and **WDBCHNT**, are reported in Appendix F.

Next, section 10.4 describes in detail the base and scope respondents' choices for the accelerated recovery program. Simple contingency tables are presented which test whether the percent that voted *for* using the **W1CH** choice measure differs across the base and scope samples as a whole and at each tax amount (W1AMT asked in W-1). Statistical tests comparing estimates of the two WTP distributions derived using the Turnbull lower-bound estimator (introduced in the previous chapter) are also presented. Section 10.4 then presents some comparisons of the two WTP distributions using parametric survival models. Finally, section 10.5 offers a summary of our findings.

§ 10.2 Design of Scope Test and Summary of Overall Findings

We are interested in testing whether WTP for the base set of injuries (WTP_B) is greater

---

As in Chapter 9, choice measures are denoted by bold capital letters.

than WTP for the scope set (WTP$_S$). Formally, this translates into a null hypothesis,

$$H_0: \ WTP_B = WTP_S \ ,$$

versus an alternative hypothesis,

$$H_1: \ WTP_B > WTP_S \ .$$

To implement this test, the selected dwelling units (DU's) were randomly assigned within segments in a 2 to 1 ratio to either the base or scope survey instrument and then to one of the five tax amount versions described in Chapter 6. This second step in the random assignment allows an evaluation of the base and scope sample responses at these different tax amounts. Further, using the same tax amounts in both the base and scope instruments facilitates comparisons using both simple contingency tables and the lower-bound estimate of the mean from the Turnbull estimator (Turnbull, 1976). The scope survey instrument was administered to a fairly large sample (final N=953) to ensure that statistical tests would have reasonable power to test the two samples' distributions of choices across tax amounts.

The most general test of sensitivity to scope is a test for differences in the distribution of the **W1CH** choices overall. We can also conduct a similar test at each of the W-1 tax amounts. Both sets of tests imply that the distributions of responses to the base and scope versions are significantly different. This is also true if we use the **W1** or **W1CHNT** choice measures.

A second type of test imposes some structure on the responses by first estimating separate

---

Twice as many base instruments were randomly assigned to selected DU's as the principal purpose of this study is to use that instrument in deriving the estimate of prospective ILUV presented in Chapter 11.

Random assignment of the cases (*i.e.*, respondents) to the two treatments (*i.e.*, base or scope) allows relatively simple and easy-to-interpret statistical techniques to be used to assess any differences in the responses to the two treatments.

See Appendix E for a detailed discussion of the Turnbull estimator.

Turnbull lower-bound means for the WTP constructed from the two sample's choices (*base*-**W1CH** and *scope*-**W1CH**). Because the samples are independent, the estimated lower bounds can be compared using a straight-forward *z-test*. This procedure tests the responsiveness to scope by testing the equality in these lower-bound means for the base and scope injuries. A third approach adds more structure by using measures of central tendency for the selected parameter from survival models to test the same hypothesis.

For all of the above tests, the conclusion remains the same — there is a clear and robust difference between the distribution of respondents' choices for the objects of choice corresponding to the base and the scope injuries. This difference leads to a significant difference in both non-parametric and parametric measures of central tendency for the distribution of WTP (across respondents). These tests are discussed in section 10.4.

§ 10.3 Examination of Other Questions in Base and Scope Samples

Given random assignment of respondents between treatments, we would not expect to see any difference in responses to three different sets of questions in the base and scope instruments:

- questions asked <u>before</u> the injuries were introduced (*e.g.,* A-1 and A-2 series),
- recreational activity and environmental questions in section B, and
- demographic questions in section C.

This expectation can be tested in a contingency table framework for each variable. The p-values for 66 variables, shown in Table F.11 of Appendix F, indicate few significant differences between the base and scope versions of the survey and indeed fewer significant differences than would be expected simply by chance at the 0.05 confidence level (given the number of comparisons, 3.3 significant differences expected versus 1 observed) or at the 0.10 confidence level (6.6 expected

versus 5 observed). The two questions with the smallest p-values (A-2d, homeless, and B-15, beach use) are not expected to be important to a scope test. We conclude from this exercise that the observed differences in the two samples are quite small and none seem likely to substantially influence any comparison of the WTP responses to the base and scope surveys.

Several questions can be used to evaluate whether there is indirect evidence that respondents understood the differences in the set of injuries described in the base and the scope surveys. Question B-8, which asked about the seriousness of the reproduction problems described, is perhaps the most likely to be influenced by whether or not the respondent was administered the base or scope instrument. The most straightforward way to detect any differences in responses to B-8 is to construct the two-by-five contingency table shown in Table 10.1. The rows depict the treatment (base or scope) and the columns represent the responses to the B-8 question after dropping the "not sure" responses. A simple $\chi^2_{(4)}$ test has a value of 148.90 and rejects the hypothesis (p < 0.001) of no difference in the B-8 responses with the version of the survey administered. That is, respondents who were administered the base version were

---

The expected number of significant differences is found by multiplying the confidence level by the number of variables compared to get the expected number of comparisons which are significantly different (*e.g.*, 0.10 x 66 = 6.6).

The base version of question B-8 was: "All things considered, would you say the *fish and bird* reproduction problems I told you about in the South Coast are not serious at all, not too serious, somewhat serious, very serious, or extremely serious?" Question B-8 in the scope version was: "All things considered, would you say the *fish* reproduction problems I told you about in the South Coast are not serious at all, not too serious, somewhat serious, very serious, or extremely serious?"

As noted in Chapter 6, the scope version of the questionnaire contained an additional question at the end of the survey which asks respondents if they thought that the reproduction problems described would be more serious if they impacted Bald Eagles and Peregrine Falcons. Seventy-four percent responded *yes*. This internal test provides additional support for the notion that respondents perceived a difference between the two injury scenarios.

significantly more likely to consider the described set of injuries as more serious compared to those who were administered the scope version.

Table 10.1 Version by Perceived Seriousness of Injury

Version	Not serious	Not too serious	Somewhat Serious	Very Serious	Extremely Serious
Base (row percent)	97 (5.3%)	355 (19.3%)	713 (38.7%)	475 (25.8%)	201 (10.9%)
Scope (row percent)	138 (14.6%)	287 (30.4%)	315 (33.4%)	140 (14.9%)	63 (6.7%)
$\chi^2_{(4)} = 148.90; p < 0.001$					

A second place we might look for indications of a difference between the response to the two versions of the survey is with respect to the length of natural recovery (questions B-2 and B-3). Due to the shorter natural recovery time in the scope version (15 years) and the longer natural recovery time in the base version (50 years), we might expect a higher fraction of the scope respondents to indicate that they thought a longer natural recovery time was likely and a higher fraction of respondents in the base version to indicate that a shorter natural recovery time was likely. These results are shown in Tables 10.2 and Table 10.3. In both instances the hypothesis of no difference is rejected (p < 0.001) in favor of the expected direction of the observed difference. There is also a significant difference (p=0.002) on B-7a for the small percent (2.0) of respondents who thought they were pushed to vote against the program — a higher fraction of those receiving

---

There is also a significant difference (p < 0.001) on the initial B-2 question with 6% more of those receiving the base instrument perceiving a different natural recovery period than stated in the survey than those receiving the scope.

the scope version felt pushed to vote against.

Table 10.2 Version by Expected *Longer* Natural Recovery Time

Version	Not Longer	Longer
Base (row percent)	1740 (93.7%)	109 (6.3%)
Scope (row percent)	860 (90.2%)	93 (9.8%)
$\chi^2_{(1)} = 10.89; p = 0.001$ Fisher's Exact Test: p = 7.36E-04		

Table 10.3 Version by Expected *Shorter* Natural Recovery Time

Version	Not Shorter	Shorter
Base (row percent)	1579 (85.0%)	278 (15.0%)
Scope (row percent)	893 (93.7%)	60 (6.3%)
$\chi^2_{(1)} = 44.79; p < 0.001$ Fisher's Exact Test: p = 1.88E-12		

There are additional questions in section B eliciting respondents' evaluations of the injuries and accelerated recovery program that are less suitable for use in discriminating respondents' perceptions of the base and scope injuries. For example, question B-1 asked: "First, did it seem to you DDT and PCB could cause the reproduction problems I told you about?" Both the base and scope surveys described reproduction problems for specific species. The key distinction between the descriptions is the number of species affected and the timing of natural recovery.

Dropping the small number of respondents who felt pushed to vote *against* in both the base and scope samples does not alter the conclusion of any of the tests of sensitivity to scope presented in this chapter and in Appendix F.

There seems no clear reason to expect these distinctions to change the way people would answer this question; our findings suggest they do not (p=0.308). Turning now to B-4, there is also no significant difference (p=0.263) on this question which asked, "When you decided how to vote, did it seem to you that the speed-up program would be completely effective in solving the reproduction problems within five years?" There is no significant difference (p=0.721) on B-6, which asked respondents: "When you decided how to vote, did you think your household would have to pay the special tax for the program for one year or for more than one year?"

The opportunities offered respondents to change their vote from a vote *for* to a vote *against* the program (see questions W-7 and C-17 to C-21 in Appendices A.1 and A.2) also afford a chance to review this issue. Here we might expect a larger fraction of respondents changing their vote in the scope version if some respondents initially perceived the scope injuries to be larger than described. Table 10.4 displays the two-by-two contingency table for those who initially voted *for* (either in W-1 or W-3) by whether they later changed their vote (in either W-7 or C-17 to C-21). These results indicate that a higher fraction of scope respondents changed their vote and that this difference is significant at any conventional confidence level.

Table 10.4 Version by Whether Respondent Changed Vote Upon Reconsideration

Version	Not Change	Change
Base (row percent)	804 (88.5%)	104 (11.5%)
Scope (row percent)	245 (81.7%)	55 (18.3%)
$\chi^2_{(1)} = 9.34$; p = 0.002 Fisher Exact Test: p = 0.003		

In the contingency table for B-5, the follow-up question to B-4, there is a significant difference but the response pattern suggests only a weak directional effect of version.

§ 10.4 Tests of the Scope Hypothesis

There are a number of different statistical tests that can be conducted to evaluate the null hypothesis, H_0: $WTP_B = WTP_S$, versus the alternative, H_1: $WTP_B > WTP_S$. The most direct test considers a two-by-two contingency table of the version of the survey by **W1CH**. As shown in the contingency table (see Table 10.5), overall 36.6% of those who were administered the base instrument voted *for* the accelerated recovery program, while 21.0% of those administered the scope version voted *for* the program. The $\chi^2_{(1)}$ test statistic, 71.10, rejects the null hypothesis ($p < 0.001$) that $WTP_B = WTP_S$.

The uniformly most powerful test of the hypothesis in a two-by-two contingency table is the Fisher's exact test. This test takes account of the direction of the inequality in the alternative hypothesis (that $WTP_B > WTP_S$) and has a p-value of less than 0.001. Thus, both the χ^2 and Fisher's exact tests indicate that respondents who were administered the base instrument were significantly more likely to vote *for* the program at the W-1 tax amount compared with those who were administered the scope instrument.

Table 10.5 Version by W1CH Aggregating Over Tax Amount Versions

Version	Not For	For

The χ^2 test for a two-by-two contingency table looks at the expected frequency in each of the four cells under the null hypothesis of no association relative to the observed frequency. This test gives the probability that the observed frequency differs from that expected under the null hypothesis. The χ^2 test in this case is a test of the null hypothesis that $WTP_B = WTP_S$ versus the alternative hypothesis that $WTP_B \neq WTP_S$ and is valid in reasonably large samples. See Lehmann (1986) for further discussion of the properties of this test.

The one-sided Fisher's exact test is more powerful because it takes into account the direction of the alternative hypothesis and it is valid for both small and large samples. See Lehmann (1986) for further discussion of the properties of this test.

Base (row percent)	1178 (63.4%)	679 (36.6%)
Scope (row percent)	753 (79.0%)	200 (21.0%)
$\chi^2_{(1)} = 71.10; p < 0.001$ Fisher's Exact Test: p = 5.81E-18		

In Table 10.5, the **W1CH** responses are aggregated over the five tax amount versions. To more closely examine how the distribution of implied WTP estimates is likely to differ between the base and scope samples, each of the tax amount versions can be examined separately using the same two-by-two contingency table.   This approach has the advantage of eliminating minor deviations due to sampling variation in the percent of respondents who were administered the base and scope versions at each tax amount. It is also a much more demanding set of tests than the one presented in Table 10.5 as it requires that the percent who voted *for* at each of the five tax amounts (W1AMT) to be larger in the base sample and, equally important, performs the test of sensitivity to scope with smaller sample sizes.

These results are displayed for each tax amount version in Tables 10.6a to 10.6e. For all five tax amounts, both the $\chi^2_{(1)}$ tests and the one-sided Fisher's exact tests clearly reject the null hypothesis at very small confidence levels. These conclusions are unchanged using either the original W-1 response (**W1**) or by a conservative treatment of respondents who do not pay California income taxes (**W1CHNT**).

Table 10.6a Version by Choice Measure (W1CH) for W1AMT=$10

Version	Not For	For
Base (row percent)	165 (44.1%)	209 (55.9%)
Scope	130	72

(row percent)	(64.4%)	(35.6%)

$\chi^2_{(1)} = 21.50; p < 0.001$
Fisher's Exact Test: p = 2.41E-06

Table 10.6b Version by Choice Measure (W1CH) for W1AMT=$25

Version	Not For	For
Base (row percent)	189 (53.7%)	163 (46.3%)
Scope (row percent)	137 (75.3%)	45 (24.7%)
$\chi^2_{(1)} = 23.50; p < 0.001$ Fisher's Exact Test: p = 6.71E-07		

Table 10.6c Version by Choice Measure (W1CH) for W1AMT=$80

Version	Not For	For
Base (row percent)	245 (67.1%)	120 (32.9%)
Scope (row percent)	161 (82.1%)	35 (17.9%)
$\chi^2_{(1)} = 14.39; p < 0.001$ Fisher's Exact Test: p = 8.07E-05		

Table 10.6d Version by Choice Measure (W1CH) for W1AMT=$140

Version	Not For	For
Base (row percent)	283 (73.5%)	102 (26.18%)
Scope (row percent)	166 (85.1%)	29 (14.9%)
$\chi^2_{(1)} = 10.00; p < 0.002$ Fisher's Exact Test: p = 8.84E-04		

251

Table 10.6e Version by Choice Measure (W1CH) for W1AMT=$215

Version	Not For	For
Base (row percent)	296 (77.7%)	85 (22.3%)
Scope (row percent)	159 (89.3%)	19 (10.7%)
$\chi^2_{(1)} = 10.85; p < 0.001$ Fisher's Exact Test: p = 5.04E-04		

Using the data from Tables 10.6a-10.6e, one can also compare the ratio of the percent who voted *for* at each W1AMT in the base sample to the percent who voted *for* in the scope sample. For the $10 tax amount, that ratio is 1.58 and generally increases over the tax amounts. The ratio is 2.08 at the highest tax amount, $215, which indicates that respondents who were administered this base version were over two times more likely to vote *for* the program relative to those who were administered the scope version.

The second class of tests that can be used to evaluate the differences in respondents' choices in the base and scope samples is one which uses these choices to estimate the Turnbull lower-bound mean for each sample and tests for differences in these estimates. This analysis is based on the first choice question and treats those changing their vote (either at W-7 or C-17 to C-21) as voting *against* the program.

The Turnbull estimator for the base-**W1CH** responses are presented in Table 10.7a and the estimator for the scope-**W1CH** responses in Table 10.7b. After computing the lower-bound estimate of the mean from each table—$63.24 for base and $34.02 for scope—we can then use the estimated standard errors for each (also displayed in Tables 10.7a and 10.7b) to construct an ordinary one-sided z-statistic. The resulting z-statistic of 7.17 rejects the null hypothesis (p <

0.001) in favor of the alternative that $WTP_B > WTP_S$. The likelihood ratio test ($\chi^2_{(5)}$=83.46), based on combining the base and scope samples, also rejects the null hypothesis (p < 0.001) of equivalent distributions.

Changing the responses of those who do not pay California taxes to votes *against* does not alter this conclusion. The lower-bound estimates of the means are $55.58 and $29.52 for the base and scope samples, respectively. The resulting z-statistic of 6.70 rejects the null hypothesis (p < 0.001) as does the likelihood ratio test. Using the double-bounded choice measures, we would again draw this same conclusion (see Tables F.12—F.14 in Appendix F).

Both the contingency table tests and the test based on the Turnbull estimator are tests based upon non-parametric estimates of the WTP distributions. It is also possible to fit parametric survival distributions to the WTP responses and test whether there is a difference between responses to the base and scope instruments. Estimates of parametric survival models with the Weibull and log-normal distributions using the single-bounded choice measures, **W1**, **W1CH**, or **W1CHNT** (see Table F.15 in Appendix F), or the double-bounded choice measures, **WDB**, **WDBCH**, or **WDBCHNT** (see Table F.16 in Appendix F), with the base and scope samples confirm the basic conclusions found with the Turnbull estimates.

Table 10.7a Turnbull Estimation Results for W1CH-*Base* Sample [N=1857]

Lower Bound of Interval	Upper Bound of Interval	Probability of Voting *For* at Upper Bound	Change in Density
$0	$10	0.559	0.441
$10	$25	0.463	0.096
$25	$80	0.329	0.134
$80	$140	0.265	0.064
$140	$215	0.223	0.042

$215	∞	0.000	0.223
Log-Likelihood -1155.65 Lower bound of estimate of mean$63.24 Standard error of estimate 2.54			

Table 10.7b Turnbull Estimation Results for W1CH-*Scope* Sample [N=953]

Lower Bound of Interval	Upper Bound of Interval	Probability of Voting *For* at Upper Bound	Change in Density
$0	$10	0.356	0.644
$10	$25	0.247	0.109
$25	$80	0.178	0.069
$80	$140	0.148	0.030
$140	$215	0.106	0.042
$215	∞	0.000	0.106
Log-Likelihood -467.78 Lower bound of estimate of mean$34.02 Standard error of estimate 2.82			

§ 10.5 Summary

The selected dwelling units for the main study were randomly assigned to one of two survey instruments—base or scope—which differed only in the descriptions of the extent of the injuries. Our analysis of the base and scope responses first considered whether questions reflecting information that should not have been influenced by the difference in the set of injuries appear to have been influenced. These findings clearly indicate they were not. Next, considering whether independent respondents recognized differences in the injuries, question B-8, asking about the seriousness of the injury, indicated that independent respondents did interpret the two injuries as reflecting different levels of seriousness.

Using a simple contingency table framework, direct tests of whether there were different responses to the choice questions in the base and scope samples leads to rejections of the null hypothesis of no difference ($WTP_B = WTP_S$) in favor of the alternative hypothesis ($WTP_B > WTP_S$). This is true for the **W1**, **W1CH**, and **W1CHNT** choice measures, aggregated across the five tax amounts and for each tax amount. That is, a higher percent of those who were administered the base instrument were willing to pay the tax amount described than those who were administered the scope instrument. This hypothesis was also tested for the three choice measures (and their double-bounded analogues) using the Turnbull lower-bound estimate for mean WTP. All of these tests reject the hypothesis of insensitivity to scope at any standard confidence level. Parametric estimates based on all three choice measures (and their double-bounded analogues) also reject the null hypothesis of scope insensitivity at any standard confidence level.

§ 11 Aggregate Estimate of Prospective Interim Lost Use Value

§ 11.1 Introduction

In this final chapter, the aggregate estimate of prospective interim lost use value (ILUV) is presented. Section 11.2 summarizes the economic and practical issues associated with defining the population to which we extrapolate our lower-bound estimate of mean willingness to pay (WTP) reported in Chapter 9. The *population* having "rights" to the injured resources in the Southern California Bight is defined using the negotiation framework introduced in Chapter 3. This theoretical definition of the population is evaluated in light of practical considerations. In Section 11.3, the weighted Turnbull estimate of the lower-bound mean is presented. Finally, Section 11.4 presents our aggregate estimate of prospective ILUV.

§ 11.2 Delineation of the *Population* for Aggregation

Chapter 3 described the theoretical definition of an individual's economic value for any object of choice. These measures of value were described as being derived from individuals' choices. It was further noted that the appropriate choice for the estimation of ILUV is one that permits the construction of *total* economic value, and an *ideal* choice would be one in which each member of the public (acting through agencies that serve as trustees) would agree to "permit" a pattern of injuries and natural recovery with restoration to a specified set of natural resources. An ideal choice defines the trade-off desired by isolating what must be given to each person (monetary compensation) for him or her to freely forego the object of choice (in this case, the resources in their baseline states).

This definition of the ideal choice for the estimation of ILUV is analogous to what would be sought if the trustees for the natural resources and those responsible for the injuries could

negotiate a payment in advance to assure that people's losses would be compensated. In undertaking a negotiation for these losses, the trustees for the resources would require estimates of total values for all individuals assumed to comprise the *public* or *relevant population* considered to have "rights" to the injured resources. Because the legal framework for recovery of ILUV does not specifically identify the relevant population, our definition is based on economic criteria as well as on the practical issues associated with the design of appropriate choices from which values can be constructed.

A definition of the relevant population specifies the set of individuals with "rights" to the injured resources and therefore could, in principle, correspond to all individuals experiencing losses due to the natural resource injuries. The compensation criteria underlying the ideal construction of ILUV would require that these individuals be in a position to have the opportunity to accept the losses (due to the injuries) in exchange for compensation specified by the trustee's negotiation. As noted in Chapters 3 and 4, this process leads to a WTA measure of economic value; however, the choice elements corresponding to a description of the object and circumstances of choice necessary for framing a WTA choice were not considered feasible. As a result, a WTP perspective on the rights to the natural resources and, consequently, a program for accelerated recovery of the injured natural resources was adopted.

These two survey design decisions have implications for the definition of the *relevant population* assumed to be represented in the negotiation framework by the trustees. A WTP choice requires that the individuals in the population accept financial responsibility for their choices (*i.e.,* in this case, a one-time increase in California income taxes). This requirement

---

The decision to adopt a WTP assignment of rights is conservative due to the wealth limitations discussed in Chapter 3.

implies that the population must be confined to adult decision-makers, *i.e.,* individuals with the ability to make such financial commitments. In principle, anyone who would experience losses under the circumstances of choice envisioned by the negotiation framework could be included in this population.

Furthermore, given the above qualification, any household in the United States could conceivably experience losses and accept financial responsibility for the program. Nonetheless, given the geographic distribution of injuries and the availability of substitutes elsewhere in the country for some of the affected resources, an individual living a substantial distance from the location of the injuries (*e.g.,* those living a substantial distance outside of California) might be expected to be willing to pay less for the program. However, on the basis of distance alone, one can not define a point of geography beyond which households would not be willing to pay *anything* for the accelerated recovery program. To identify those who would and would not be willing to pay for the program is an empirical issue.

The estimation of ILUV requires that respondents be offered a credible choice. Following the NOAA Panel recommendations and our own past experience, the choice presented in our study was framed in a discrete take-it-or-leave-it referendum with the tax amount described as a one-time, state tax payment. The relevant population was confined to California households. As noted above, it is reasonable to assume that households outside of California could have a positive willingness to pay for the accelerated recovery program offered in this study. However, expanding our population definition to areas outside of California, for example Arizona, Oregon, and Washington, would require a very different characterization of the choice. For example, while it is credible to respondents residing in California to tax themselves to mitigate injuries to

---

Nothing in economic theory *requires* this presumption to be true.

what they may perceive as state resources, it is unclear whether residents of Washington would find it credible to be taxed to "fix" what might be perceived by some as California problems.

While it is possible to design a choice that would be credible to residents outside of California, such design work would take a great deal of time and effort. The time constraints imposed by the damage assessment suggested that the prudent strategy would be to limit the relevant population to California households. To the extent people elsewhere in the U.S. experience losses and would be willing to pay for the program, this decision understates the aggregate estimate for prospective ILUV.

Within the population of California households it was necessary to further limit the population to which the survey was design to be extrapolated. First, as described in Chapter 5, similar time constraints led us to conduct interviews only in English, thereby excluding an estimated 697 thousand non-English speaking households from the inference population. Additionally, as described in Chapter 7, given the close proximity of 35 selected dwelling units to a neighborhood where the excavation of DDT was causing intense concern about human health, we excluded the 63 thousand households in the surrounding Torrance area (defined by five zipcodes). Thus, our results may be extrapolated to the population of English-speaking households in California outside of this Torrance area. To the degree that non-English speaking households in California or those in the Torrance area would be willing to pay for the program, we will underestimate prospective ILUV.

Finally, it should be noted that we make a very conservative adjustment with respect to California households that do not pay state income taxes. The use of a one-time increase in California income taxes as one of the choice elements implies that individuals who are not eligible to pay California income taxes may respond differently to the choice question than those who do

pay taxes. This was borne out by the multivariate analysis reported in Table 9.7 which showed that households not paying California income taxes last year were more likely to support the program, holding all other aspects of their characteristics and the tax amount (W1AMT) constant. Given the likelihood that some of these respondents assumed that they would not have to accept financial responsibility for their choices, these respondents' *for* votes have been treated as *not for* votes for the purposes of developing the lower-bound estimate of aggregate WTP.   This effectively treats this group of respondents as having a lower-bound mean willingness-to-pay of zero. It is likely that this group of respondents has a positive, not a zero, aggregate (for this group) prospective ILUV for the program. The decisions involving these groups were intended to avoid any possible upward bias in the lower-bound estimate of mean WTP as a summary statistic for the population to which we extrapolate.

§ 11.3 Population Estimate of the Turnbull Lower-Bound Mean WTP

As described in Section 7.10, samples will vary somewhat from the populations from which they are drawn. As a result, it is standard survey practice to apply sample weights in order to estimate *population* values. In this section, we present the population estimate of the lower-bound mean using the weights developed by Westat (see Appendix C.5.5 for details of how the weights were constructed).

Table 11.11 provides the weighted estimate of the Turnbull likelihood function for **W1CHNT**. The population lower-bound mean estimate of $55.61 is only $0.03 higher than that displayed for the **W1CHNT** sample estimate in Table 9.9. The standard error of $2.66 in Table

---

See Appendix E.

11.1 was computed taking into account the particular sample design used in our study.

Table 11.1 Weighted Turnbull Estimate of WTP Distribution and Lower-Bound Mean: Using W1CHNT Choice Measure [N=1857]

Lower Bound of Interval	Upper Bound of Interval	Probability of Voting *For* at Upper Bound	Change in Density
$0	$10	0.489	0.511
$10	$25	0.399	0.090
$25	$80	0.289	0.110
$80	$140	0.237	0.052
$140	$215	0.194	0.043
$215	∞	0.000	0.194

Log-Likelihood -1113.39
Estimate of lower-bound mean $55.61
Jackknifed Standard error of the estimate $2.66

The fact that the weights produce little difference is likely due to several considerations: (1) the divergence between the sample and the population on the demographic and locational characteristics on which the weights were based was generally small, (2) placing all of the density

---

Tables F.17, F.18, and F.19 contain the population lower-bound mean estimates respectively for the **W1CH**, **WDBCH**, and **WDBCHNT** choice measures. These estimates are also quite close to their counterparts reported earlier in Chapter 9 and Appendix F.

As noted in Chapter 7, our sample design involved both clustering (at the final stage of selection) and the construction of sample weights. Relative to a simple random sample, both clustering and weighting decrease efficiency in the sense of needing more observations for the same level of precision. In order to estimate the standard error of the lower-bound mean estimate in Table 11.1, we used a resampling technique known as the jackknife, which takes account of these complexities in the sampling design (Wolter, 1985). Effectively this approach simulates what would happen if one drew repeated samples from the population using the same complex sampling design. The creation of these jackknife replicate weights is described in Appendix C.5.5. The jackknifed standard error is 9.7% larger than the standard error derived under the assumption of simple random sampling.

261

in the Turnbull estimator at the lower bound of each interval makes the lower-bound estimate of the mean fairly robust to the small changes the weights represent to the composition of the sample, and (3) setting all non-taxpayers to *not for* votes further decreases the sensitivity of the estimate to weighting the Turnbull likelihood function.

§ 11.4 Aggregate Estimate of Prospective ILUV

The aggregate estimate of prospective ILUV is obtained in a straightforward manner by multiplying the W1CHNT, lower-bound Turnbull estimate of the mean for the population, $55.61, by the estimate of the number of households in the population (10,347,108) to which the survey was designed to extrapolate. This extrapolation yields a point estimate of $575,402,676 with a standard error of $27,523,307.

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*, 58, 4601-4614.

BAILEY, K. D. (1987): *Methods of Social Research*, third edition, New York: Free Press.

BOCKSTAEL, N. E., K. E. MCCONNELL AND I. STRAND (1991): "Recreation," in *Measuring the Demand for Environmental Commodities*, ed. by J. B. Braden and C. Kolstad, Amsterdam: North-Holland, 227-270.

CARP, F. (1974): "Position Effects on Interview Responses," *Journal of Gerontology*, 29, 581-587.

CARSON, R. T., N. E. FLORES, K. M. MARTIN AND J. L. WRIGHT (1994): "Contingent Valuation and Revealed Preference Methodologies: Comparing the Estimates for Quasi-Public Goods," Discussion paper 94-7, May, Department of Economics, University of California, San Diego.

CARSON, R. T., R. C. MITCHELL, W. M. HANEMANN, R. J. KOPP, S. PRESSER AND P. A. RUUD (1992): *A Contingent Valuation Study of Lost Passive Use Values Resulting from the Exxon Valdez Oil Spill*, Report to the Attorney General of the State of Alaska, Volumes I & II, Natural Resource Damage Assessment, Inc., November.

CARSON, R. T. AND R. C. MITCHELL (Forthcoming): "Sequencing and Nesting in Contingent Valuation Surveys," *Journal of Environmental Economics and Management*.

CARSON, R. T., J. L. WRIGHT, N. J. CARSON, A. ALBERINI AND N. E. FLORES (1994): *A Bibliography of Contingent Valuation Studies and Papers*, La Jolla, CA: Natural Resource Damage Assessment, Inc., March.

CIRANCY-WANTRUP, S.V. (1947): "Capital Returns from Soil-Conservation Practices," *Journal of Farm Economics*, 29, 1181-1196.

CONVERSE, P.E. (1964): "The Nature of Belief Systems in Mass Publics," in *Ideology and Discontent*, ed. by D. Apter, Glencoe, IL: Free Press, 206-261.

COUNCIL OF AMERICAN SURVEY RESEARCH ORGANIZATION (1982): *On the Definition of Response Rates*, Report of the CASRO Task Force on Completion Rates, Port Jefferson, NY, June.

DAILEY, M. D., D. J. REISH AND J. W. ANDERSON (1993): *Ecology of the Southern California Bight: A Synthesis and Interpretation*, University of California Press.

DAVIS, R. K. (1963): *The Value of Outdoor Recreation: An Economic Study of the Maine Woods*, Dissertation, Harvard University.

DESVOUSGES, W. H., V. K. SMITH, D. H. BROWN AND D. K. PATE (1984): *The Role of Focus Groups in Designing a Contingent Valuation Survey to Measure the Benefits of Hazardous Waste Management Regulations*, Draft report for the U.S. Environmental Protection Agency, Research Triangle Park, N.C.: Research Triangle Institute.

DUFFIELD, J. W. AND D. A. PATTERSON (1991): "Field Testing Existence Values: An Instream Flow Trust Fund for Montana Rivers," Paper presented at the annual meeting of the American Economic Association, New Orleans, January.

GREENBAUM, T. L. (1993): *The Practical Handbook and Guide to Focus Group Research*, second edition, Lexington, MA: Lexington Books.

GRICE, H.P. (1975): "Logic and Conversation," in *Syntax and Semantics 3: Speech Acts*, New York: Academic Press, 41-58.

HANEMANN, W.M. (1991): "Willingness to Pay and Willingness to Accept: How Much Can   They Differ?

HANEMANN, W. M., J. B. LOOMIS AND B. KANNINEN (1991): "Statistical Efficiency of Double-Bounded Dichotomous Choice Contingent Valuation," *American Journal of Agricultural Economics*, 73(4), 1255-1263.

HASTIE, T. AND R. TIBSHIRANI (1990): *Generalized Additive Models*, New York: Chapman and Hall.

HICKS, J. (1943): "The Four Consumer Surpluses," *Review of Economic Studies*, 11(1), 31-41.

------ (1939): "The Foundation of Welfare Economics," *Economics Journal*, 49, 696-712.

HOUTHAKKER, H. (1950): "Revealed Preference and the Utility Function," *Economica*, NS17, 159-174.

JOBE, J. AND D. MINGAY (1989): "Cognitive Research Improves Questionnaires," *American Journal of Public Health*, 79, 1053-1055.

KEMP, M. A. AND C. MAXWELL (1992): "Exploring a Budget Context for Contingent Valuation Estimates," Paper presented at the Cambridge Economics, Inc. Symposium, Contingent Valuation: A Critical Assessment, Washington, DC, April.

KISH, L. (1965): *Survey Sampling*, New York: John Wiley.

KNETSCH, J. L. AND R. K. DAVIS (1966): "Comparisons of Methods for Resource Evaluation," in *Water Research*, ed. by A. V. Kneese and S. C. Smith, Baltimore, MD: Johns Hopkins University Press for Resources for the Future, 125-142.

KREPS, D. M. (1990): *A Course in Microeconomic Theory*, Princeton, NJ: Princeton University Press.

KROSNICK, J.A. (1991): "Response Strategies for Coping with the Cognitive Demands of Attitude Measures in Surveys," *Applied Cognitive Psychology*, 5, 213-236.

KROSNICK, J. A. AND D. F. ALWIN (1987): "An Evaluation of a Cognitive Theory of Response-Order Effects in Survey Measurement," *Public Opinion Quarterly*, 51, 201-219.

KROSNICK, J. A., F. LI AND D. R. LEHMAN (1990): "Conversational Conventions, Order of Information Acquisition, and the Effect of Base Rates and Individuating Information on Social Judgments," *Journal of Personality and Social Psychology*, 59(6), 1140-1152.

LEHMANN, E.L. (1986): *Testing Statistical Hypotheses*, second edition, New York: John Wiley and Sons.

LITTLE, I.M.D. (1949): "A Reformation of the Theory of Consumers' Behavior," *Oxford Economic Papers*, NS1, 90-99.

LODGE, M., K. M. MCGRAW AND P. STROH (1989): "An Impression-Driven Model of Candidate Evaluation," *American Political Science Review*, 83(2), 397-419.

LOS ANGELES TIMES, March 10, 1994, at B3.

MITCHELL, R. C. AND R. T. CARSON (1989): *Using Surveys to Value Public Goods: The Contingent Valuation Method*, Washington, DC: Resources for the Future.

NATIONAL OPINION RESEARCH CENTER (1993): *National Issues Study (NIS) Data Collection and Sampling Final Report*, NORC, University of Chicago, October.

NATURAL RESOURCE DAMAGE ASSESSMENT, Inc. (1994): *Ballot Box and No-Vote Option Draft Report*, Report prepared for the National Oceanic and Atmospheric Administration, Volumes I & II, La Jolla, CA, August.

NISBETT, R.E. AND T.D. WILSON (1977): "Telling More than We Know: Verbal Reports on Mental Processes," *Psychological Review*, 84, 231-259.

PORTNEY, P. (1994): "The Contingent Valuation Debate: Why Economists Should Care," *Journal of Economic Perspectives*, 8, no. 4.

PREGIBON, D. (1982): "Resistant Fits for Some Commonly Used Logistic Models with Medical Applications," *Biometrics*, 38, 485-498.

RANDALL, A. (1994): "The Travel Cost Method," *Land Economics*, 70(1), 88-96.

RANDALL, A., B. IVES AND E. EASTMAN (1974): "Benefits of Abating Aesthetic Environmental Damage from the Four Corners Power Plant, Fruitland, New Mexico," Agricultural Experiment Station Bulletin No. 618, New Mexico State University, May.

SAMUELSON, P. A. (1938): "A Note on the Pure Theory of Consumer's Behaviour: An Addendum," *Economica*, 5, 353-354.

------ (1948): "Consumption Theory in Terms of Revealed Preference," *Economica*, 15, 243-253.

------ (1938): "A Note on the Pure Theory of Consumer's Behaviour," *Economica*, 5, 61-71.

SCHUMAN, H. AND S. PRESSER (1981): *Questions and Answers in Attitude Surveys*, New York: Academic Press.

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(1), 91-106.

TETLOCK, P. (1983): "Accountability and the Perseverance of First Impressions," *Social Psychology Quarterly*, 45, 74-83.

TURNBULL, B. (1976): "The Empirical Distribution Function with Arbitrary Grouped, Censored, and Truncated Data," *Journal of the Royal Statistical Society*, 38, 290-295, Series B.

TURNER, C. F. AND E. MARTIN (1984): *Surveying Subjective Phenomena*, New York: Russell Sage Foundation.

UNITED STATES. BUREAU OF THE CENSUS (1993): *Money Income of Households, Families, and Persons in the United States: 1992*, Current Population Report, Series P60-184, Washington, DC.

UNITED STATES. COURT OF APPEALS FOR THE DISTRICT OF COLUMBIA: "Ohio v. United States Department of Interior," 880 F.2nd 432 (D.C. Cir. 1989).

UNITED STATES. WATER RESOURCES COUNCIL (1979): "Procedures for Evaluation of National Economic Development (NED): Benefits and Costs in Water Resources Planning (Level C), Final Rule," *Federal Register*, 44, 72892-72976.

VARIAN, H. R. (1992): *Microeconomic Analysis*, third edition, New York: WW North, 132-133, 138-139.

WESTAT, Inc. (1994a): *California Issues Study: Interviewer's Manual - Main Study*, Rockville, MD, March.

------ (1994b): *California Issues Study: Main Study Trainer's Guide*, Rockville, MD, February.

WILLIG, R. (1976): "Consumer's Surplus Without Apology," *American Economic Review*, 66, 589-597.

WOLFINGERS, R. E. AND S. J. ROSENSTONE (1980): *Who Votes?*, New Haven: Yale University Press.

WOLTER, K. M. (1985): *Introduction to Variance Estimation*, New York: Springer-Verlag.

SIGNATURE PAGES