Improving Accuracy and Reducing Cost of Environmental Benefit Assessments

Field and Laboratory Experiments On the Reliability of the Contingent Valuation Method

by

William D. Schulze
Gary H. McClelland
David J. Schenk
Julie R. Irwin
Steven R. Elliott

Rebecca R. Boyce
University of Colorado at Boulder
Boulder, Colorado

Thomas Stewart
State University at New York, Albany

Paul Slovic
Sarah Lichtenstein
Decision Research, Eugene, Oregon

Leland Deck
USEPA-OAQPS Research Triangle Park, N.C.

Mark Thayer
California State University at San Diego

Prepared by the University of Colorado under
U.S.E.P.A. Cooperative Agreement #CR-812054

September 1993

Project Officer
Dr. Alan Carlin
Office of Policy, Planning and Evaluation
U.S. Environmental Protection Agency
Washington, D.C. 20460

DISCLAIMER

TABLE OF CONTENTS

CHAPTER 1

Survey Values for Public Goods: A Field Test of the Contingent Valuation Method

1.1 Introduction

Use of surveys for valuing public goods became commonplace in the decade of the 1980's. The contingent valuation method (CVM) has, as a result of the need for damage assessments under the Comprehensive Environmental Response, Compensation and Liability Act of 1980 (CERCLA), now undergone Federal Court review. The decision of the court supports use of the method as a legitimate alternative to property value or other hedonic methods and to the travel cost approach in measuring natural resource damages. In great part, acceptance of the method has been based on a series of studies in which value estimates obtained by asking respondents for their willingness to pay (WTP) were compared to values obtained from indirect approaches such as the hedonic or travel cost method (see, for example Brookshire, Thayer, Schulze, d'Arge, 1982; or Smith, Desvousges and Fisher, 1986). In several field experiments actual purchase decisions have been compared to hypothetical purchase decisions (Bishop and Heberlein, 1978 and Dickie, Fisher, and Gerking, 1987). In all of these studies hypothetical behavior was sufficiently predictive of actual behavior that researchers concluded meaningful values could be obtained for benefit-cost analysis.

However, in their extensive review of the CVM literature, Cummings, Brookshire and Schulze (1986) note that in all of the available comparison studies, respondents necessarily had obtained at least some market experience with the commodity. For example, Brookshire, et. al. (1982) compared survey values obtained for air quality improvement in the Los Angeles area with air quality values obtained from a property value study. The premium found in the home sale market for areas with cleaner air is well known by area residents who experience a trade-off between housing costs and air quality in choosing where to live. Thus, Cummings, Brookshire and Schulze argue that direct evidence has not been obtained on respondents' abilities to provide meaningful hypothetical values for public goods for which respondents have little or no prior market experience. By their very nature, many public goods do not allow market experience. For example many studies using the CVM have shown large existence and bequest values for preserving environmental commodities (e.g., Greenley, Walsh and Young, 1981 and Schulze et. al. 1983). Freeman (1987) has argued that only the CVM can be used to measure those values since such preferences are not reflected in existing markets, denying both market experience and preventing use of indirect methods for valuation. Hypothetical bias, defined as the difference between the distribution of hypothetical bids obtained from a survey and the distribution of bids that would obtain in a real world incentive compatible market setting, is the central issue in application of the CVM. Both lack of market experience and details of survey design may contribute to hypothetical bias.

To estimate hypothetical bias when respondents lack prior market experience, researchers have turned to laboratory experimental methods. Laboratory experiments typically place subjects in an unfamiliar environment (either with respect to the commodity, the market, or both) and compare hypothetical responses to actual laboratory market responses where repeated trials are used to provide market experience (e.g., Coursey, Hovis, Schulze, 1987). The next section briefly summarizes what has been learned from such experiments and, drawing on these experiments, proposes both a specific model of hypothetical bias and suggests an econometric approach for analysis of contingent values which may reduce hypothetical bias.

The impact of survey design on hypothetical bias is captured in the psychological notion of context where it is argued that apparently innocuous changes in wording, ordering, or form of questions can, under some circumstances, have a large impact on hypothetical responses. Schuman and Presser, in their study of the effect of context on survey responses (1981), argue that context effects are greatly reduced when attitudes have become crystallized. We discuss both context and crystallization in the next section and propose a conceptual framework and a strategy for dealing with possible context effects in obtaining contingent values.

The overall objective of this study is to test our proposed procedures for dealing with hypothetical bias in a field application of the CVM. The commodity chosen for the study, air quality in the Denver metropolitan area, has three features which make it appropriate for such a test. First, a careful psychological study of how residents perceive air pollution in the region is available (Stewart et al. 1983, 1984). Second, one of the primary features of Denver's air pollution problem, the "Brown Cloud" which obscures views of both the center city skyline and of the Colorado Front Range, is visible throughout the city. Thus, property value markets are little affected by air pollution, so residents have had little or no market experience with the commodity. Third, a high level of awareness of the problem and a community consensus that something must be done has been achieved in the region. For example, the Chamber of Commerce has strongly supported new proposed air pollution controls and such innovative measures as currently required use of oxygenated fuels have received wide public support. Thus, although residents have had little or no market experience with the commodity, most have at least thought about the problem. Our choice of commodity can thus be seen as an attempt to move away from market experience while still retaining a commodity for which the public has a clear sense of both the nature and importance of the commodity itself.

The remainder of the chapter is organized as follows: Section 1.2 discusses hypothetical bias. The design of the survey instruments used to value air quality is presented in Section 1.3. Section 1.4 presents results and data analysis and Section 1.5 provides conclusions and recommendations for future research, and outlines the remainder of the report.

1.2 Evidence on Hypothetical Bias

Two primary issues remain in the application of the CVM. First, the importance of market experience with a commodity has proven difficult to quantify using surveys alone. Survey values obtained in the field tend to be bimodally distributed with a large number of zero bids and an upper mode which is skewed, showing a thick tail of high bids. Researchers have tended to view both the large number of zero and very high bids with considerable skepticism. Fortunately, laboratory experiments have shed considerable light on this problem. Second, if context effects are important, that is, if different survey designs obtain very different values for the same commodity, then estimated values are not robust. The two issues are obviously related in that if context has a large impact on values, then hypothetical bias is likely severe - especially in cases where seemingly trivial or irrelevant changes in context have a large impact on values. We consider each of these issues in turn and suggest a specific set of procedures for each.

Market Experience

Results from laboratory experiments show a consistent and striking pattern. Hypothetical bids obtained from subjects for a commodity show an increased variance relative to bids obtained in a laboratory market. Further, increasing market experience (repeated rounds in a particular auction institution) and increasing incentives (increased payoffs for participation in a particular market institution) both tend to reduce variance in bidding. The first experiment to compare hypothetical bids to auction behavior, undertaken by Coursey, Hovis and Schulze (1987), used a bitter tasting liquid, sucrose octa acetate, which was unfamiliar to subjects as the commodity. Subjects were first given a careful description of the commodity and then were asked how much they would pay to avoid a taste experience. Second, subjects were allowed to taste the liquid prior to being asked again for their willingness to pay (WTP). In this second stage subjects were familiar with the commodity but had no market experience. Third, subjects participated in a competitive auction submitting bids to avoid the commodity. Mean bids (variance) were as follows: Hypothetical with no experience $2.60 ($15.80); hypothetical with experience with the commodity $2.27 ($5.06); and actual auction bids with market experience $1.95 ($5.23). Note, the variance is much greater for the inexperienced hypothetical bids. However it appears that the decrease in variance was associated with experience with the commodity rather than with experience with the market institution. Other recent experiments which allowed more rounds of actual market experience than the Coursey, Hovis and Schulze experiment show a continued decline in bidding variance both with market experience and reward size (see Irwin, McClelland and Schulze, 1989 and Cox, Smith and Walker, 1989). Figure 1 shows how increasing variance in hypothetical bidding can bias estimates of actual behavior. First, in the top panel, if we assume that the negative values of the left tail of the hypothetical bid distribution produce zero bids (which are more common in hypothetical bidding than in actual bidding in auction experiments), then the mean bid of the hypothetical bid distribution will overestimate the mean of the actual bid distribution. This situation explains much of the decrease in mean bid in the Coursey, Hovis, and Schulze experiment. Second, in the middle panel of Figure 1 a skewed hypothetical distribution, relative to the actual bidding distribution is shown. Here the extended right hand tail is the source of an upward bias in the mean hypothetical bid. This source of error dominates the results of the experiment conducted by Irwin, McClelland, and Schulze (1989). The bottom panel shows a situation where an increase in hypothetical bidding variance produces an unbiased estimate of mean bid as occurred in the experiments conducted by Cox, Smith and Walker (1989) which varied experimental inducements.

Given the experimental evidence summarized above, what model can be used to explain hypothetical bias that might result in field surveys from a lack of market experience? Assume for simplicity that all individuals have the same true willingness to pay, W. However, the bid they provide in response to a hypothetical question about willingness to pay is B. Then where e is hypothetical error with a frequency distribution $f(\epsilon)$,

$$B = \begin{cases} W + \epsilon & \text{for } W + \epsilon > 0 \\ 0 & \text{for } W + \epsilon \leq 0 \end{cases} \qquad (1)$$

since negative values of $W + \epsilon$ imply corner solutions in the consumer choice problem. The first problem in this simple situation is, that the mean of B,

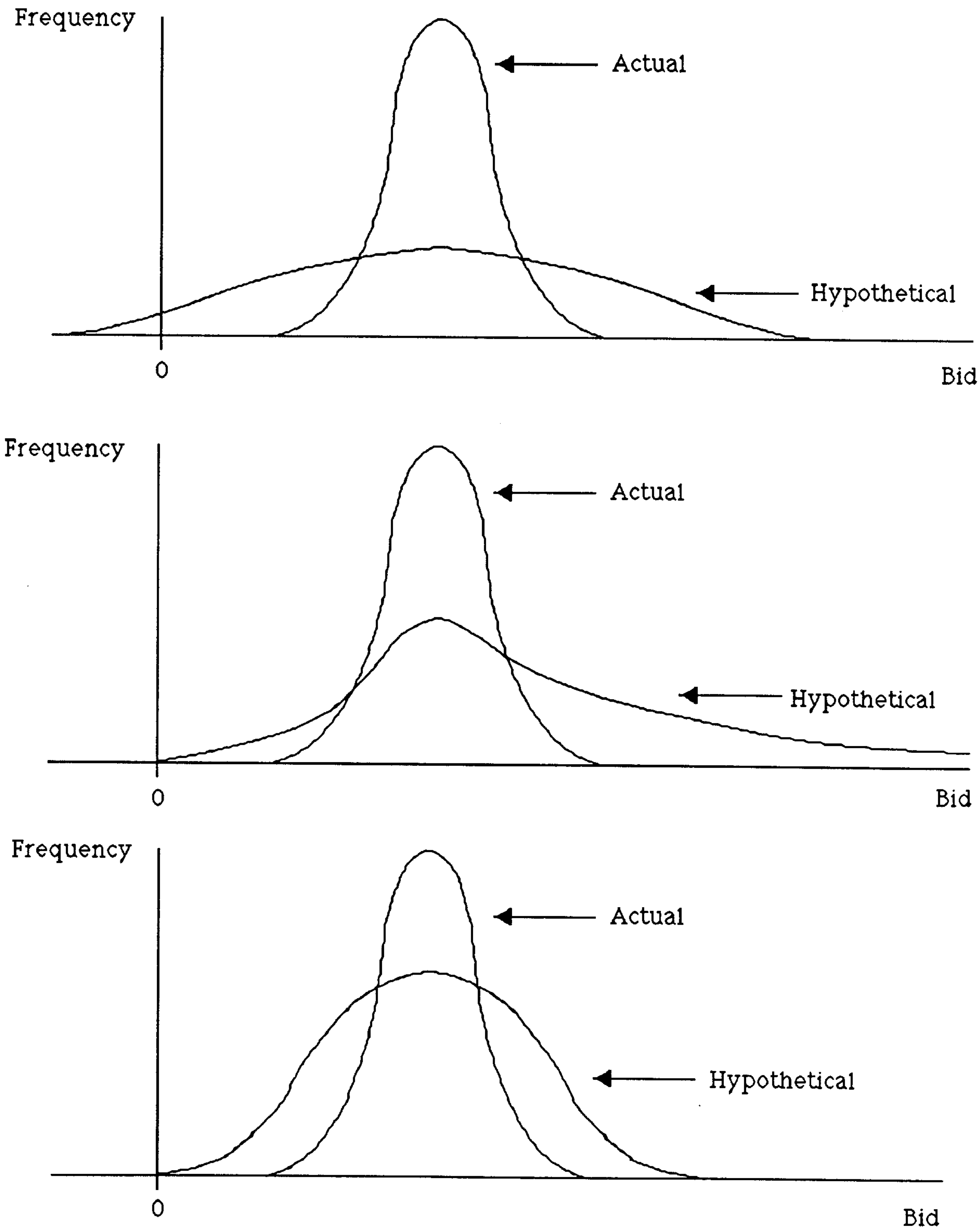
$$\bar{B} = \int_{-W}^{\infty} f(\varepsilon)\,(W+\varepsilon)d\varepsilon \neq W$$

is a biased estimate of W because of truncation bias even if we assume that hypothetical error is symmetrically distributed so that

$$\int_{-\infty}^{+\infty} f(\varepsilon)\,\varepsilon d\varepsilon = 0. \qquad (2)$$

In this situation use of the mean bid to estimate WTP will typically overstate the true value. Similarly regression estimates used to predict WTP will be upwardly biased unless Tobit is used to account for truncation (Madala, 1983).

Figure 1


Frequency
Actual
Hypothetical
0
Bid
Frequency
Actual
Hypothetical
0
Bid
Frequency
Actual
Hypothetical
0
Bid

Second, $f(\epsilon)$ may not be symmetrically distributed. As noted above, available data both from the field and laboratory suggest that $f(\epsilon)$ is often skewed. Assume for purposes of illustration that ϵ is distributed log normally so

$$\ln B = \ln W + \epsilon.$$

The simple mean $\bar{B}$, will be upward biased, but the mean of lnB will be an unbiased estimator of lnW since (2) still holds. This implies that a transformation of survey bids (such as the natural logarithm) which produces a symmetrical error distribution may well reduce hypothetical bias by eliminating skew. We propose use of a more general transformation, the Box-Cox, $(B^{\alpha}-1)/\alpha$, where α is determined to effectively normalize the error distribution in regression analysis (Box and Cox, 1964). It incorporates both the linear ($\alpha = 1$) and natural logarithm ($\alpha = 0$) transformations as possibilities. Use of this procedure has several advantages. In the past large suspect bids obtained in the CVM have been removed through trimming (e.g., Desvousges, Smith and Fisher, 1987). Trimming procedures remove large outliers which deviate from an estimated linear regression model by exceeding some predetermined statistical threshold. However, in the situation where the bid distribution shows a thick upper tail, the mean of predicted bids falls as that threshold is lowered, making final estimated values dependent on the threshold chosen. If skew is present, the procedure we propose will also lower mean values if bids generated by the estimated regression equation are used in calculating the mean. However, the reduction in predicted mean bid will be determined by the estimated value of α, the Box-Cox parameter, so as to make $f(\epsilon)$ as normal as possible. If hypothetical error dominates the residual, then it is obviously desirable to use an estimating procedure which does not bias the estimated coefficients through a skewed hypothetical error distribution.

A final problem in the interpretation and analysis of contingent values, which we feel is unresolved, is the presence of protest zero bids or refusals to bid when respondents are asked for WTP. When pretesting survey instruments, researchers in debriefing respondents have often found that zero bids are not associated with a zero value to the respondent, but rather the respondent does not feel responsible for the problem. For example, a respondent may argue that although she is harmed by air pollution, she is not responsible for creating the problem (e.g., she does not own a car). Rather, industry and others should pay and are morally responsible for cleaning up the problem. Interestingly, such respondents when asked for their willingness to accept to allow a decrease in environmental quality often refuse any amount of money, arguing that to do so would be morally wrong. Thus, moral reasoning results in an unwillingness of respondents to provide any tradeoff between money and the public good in question. An L-shaped indifference curve between money on the vertical axis and the public good on the horizontal axis results from the application of moral reasoning, a situation similar to that described by Hahneman (1989). As argued by Smith and Desvousges (1987), the absence of positive bids from such respondents results in a potential selection bias problem since as many as one third of respondents may refuse to provide values. In estimating a regression model for those respondents who do provide a WTP value, selection bias must be accounted for to obtain unbiased coefficients (Heckman, 1979). However, a question remains as to whether or not to include predicted values from the resulting estimated equation for those individuals who refused to bid. We propose that mean bid should be calculated two ways, first assuming zero values for those who refuse to bid and second using predicted values for these respondents. Note that identification of bid refusals can be accomplished in at least three ways in the design of a survey instrument. First, a question asking why the respondent bid zero can be included. Second, a question asking for willingness to accept for a decrease in the level of provision can be included along with the WTP question. If the respondent indicates that no amount of money is morally acceptable as compensation for a reduction in provision, an associated zero WTP can be rejected as inconsistent. Third, questions asking how concerned about, bothered by, or important the commodity is to the respondent can be used to check for the consistency of a zero bid.

Context

A potential problem in the design of any survey instrument is the degree to which the wording of the survey can affect respondents' answers. Both the wording of the valuing question and the information surrounding the question, which we term the *context* of the question, can affect the value given. Hogarth (1982) in an edited volume presents a number of papers that confirm the notion that context can affect people's responses, even in situations in which the context should logically have no effect. For example, researchers (e.g., Noell-Neumann, 1970) have found that the order in which (independent) questions are asked can affect people's answers to the questions. Other researchers (e.g., Tversky and Kahneman, 1986, Lichtenstein and Slovic, 1971, 1973) have found that how the question is expressed (e.g., in terms of losses versus gains, or in terms of percentages instead of fractions) can affect people's responses.

In order to understand context effects, it is helpful to think of values as being more or less *crystallized* (Schuman and Presser, 1981). If a person has had the opportunity to think about and/or declare a value for a commodity (in a marketplace, for example) to such an extent that the value is "set" in the mind, then it is unlikely that the manner in which the value is elicited will affect the value. In such a case, we would say that the commodity's value is *crystallized*, and relatively impervious to context effects. If, on the other hand, the commodity is one for which the person does not have a set value, because the commodity is not traded in a marketplace and the person has not thought of the commodity in monetary terms, then the value for that commodity is less crystallized. In such a case, context difference could affect the value in two ways.

One way in which context can have an effect is in the process of evaluating the commodity. For example, when a respondent reads the words “air quality”, many components of the concept “air quality” may come to mind. In a sense, the words “air quality” themselves may have different meanings for different people, and for many people, the meanings will be quite vague and unformed. Context can help clarify the concept of air quality, or place emphasis on different aspects of the problem. Evidence has shown (Tversky and Kahneman, 1974) that whatever components of a commodity are most cognitively available to the respondent will figure most strongly in the evaluation of the commodity. For example, reading information about the health effects of particulates in the air in Denver could cause respondents to place more emphasis on the health dangers of air pollution, thus raising their values for improving air quality. It is as if the context provides an anchor, a way to think about the problem, for the respondent.

A specific difficulty associated with the way in which values are obtained concerns the hedonic nature of many public commodities. For example, air quality improvement provides value both from reduced health effects and from improved visibility. Fischhoff and Furby (1987) suggest that people may not be able to provide component values in such cases because they have only considered air quality as a whole. Mitchell and Carson (1989), based on pretesting with focus groups, argue that people treat health improvements and visibility improvements as inseparably related. If this is the case, even if the physical situation allows health and visibility to be improved separately by controlling different pollutants, respondents will provide values as if they are joint products. If attempts are then made to separately value one component, respondents will provide a value for both components since they assume that components of value move together. Values obtained separately will add up to more than that obtained by directly asking respondents for the total value of air quality improvement. Thus, a superadditivity problem will exist. One way to overcome this difficulty is to ask respondents to split total value between health and visibility after they have given a total value. This approach is consistent with their “maintained view” that these are joint products. Unfortunately, however, superadditivity may be a more pervasive problem in that, if respondents have not previously had experience in breaking down values, stated components of value may not generally add up in a logically consistent way. It is important to test for logical consistency by varying context to check for superadditivity problems. If such a problem exists then the survey can be designed to assist respondents in deriving logically consistent answers (e.g., percent splits of a total value).

Another type of context effect is rooted in the difficulty that people have in assigning values to commodities such as *air quality*. Context can help respondents understand how the general concept of “air quality” can be translated to a monetary scale, especially since it is likely that many respondents have not thought of environmental commodities in monetary terms. It is important, for that reason, that CVM questions be given enough context so that respondents believe and understand that their money would actually buy the commodity they are evaluating.

Some researchers (Fischhoff and Furby, 1989) have suggested that even seemingly minor wording differences in CVM questions could result in respondents valuing essentially different commodities, thus making interpretation of the results impossible. The disadvantage to giving too much contextual information of this kind is the danger that the respondents will have a response to the contextual information that is unrelated to their actual values for the commodity. For instance, asking respondents if they would be willing to be "taxed" to pay for "governmental programs" that would clean up the air pollution in Denver may result in angry refusals to give a value, not because the air pollution problem is unimportant to the respondents, but because of disgust for the government, or taxes, or some other element of the questions' context.

In order to test the degree to which values are crystallized, as well as the degree to which respondents need contextual information to make sense of CVM questions, we propose that context be varied across different versions of the survey. If values obtained across survey versions are robust with respect to changes in context, then it is likely that values are crystallized and hypothetical bias may not be a severe problem for the particular commodity.

1.3 Survey Design and Variables

This section contains descriptions of the general design of the eight survey versions (see Table I for a summary of the survey designs), first outlining the common elements of the eight versions, and then listing and describing the elements that varied across survey versions.

The introductory sections of the surveys were similar across version: respondents indicated how bothered they were by the Denver Brown Cloud (the common local name for the air quality problem in Denver), where they had heard about the Brown Cloud problem, and so on. Following this set of questions, each survey asked respondents to look at a color photo insert. The insert contained 6 pictures of the Denver area under different visibility conditions and a seven-step ladder with two example photos anchoring steps 2 and 6. The photos used to anchor the ladder had been previously rated in pre-testing. Respondents used the 1-7 scale to rate each of the six example photos, first in terms of visibility, and then in terms of healthiness of the air.

The next section asked respondents to think about the current overall air quality on an average winter day in the Denver metro area and to provide a rating for this average day (winter is the poorest season for air quality in Denver). The last section of the survey, the demographic section, was the same for all versions. In this section, respondents were asked their age, income level, and so on.

Table 1

Summary of Brown Cloud Survey Design Features

		BASE		THREE		VOTING	FREQ. DIST.	3 COMMODITY COMPARISON	
	VERSION	A	B	C	D	E	F	G (WTP)	H (Choice)
RESPONSE FRAME	WTP	X	X	X	X	X	X	X	
	WTA	X	X			X			
HEALTH vs. VISIBILITY	3 Questions			X	X				
	% Split	X	X						
FORM OF THE VALUE QUESTIONS	Std. CVM	X	X	X	X		X	X	
	Voting					X			
	Choice								X
DESCRIPTION OF CHANGE IN AIR QULAITY	Average Air Quality Change	X	X	X	X	X		X	
	Freq. Distribution of Air Qual. Change						X		
CONTEXT/ INFORMATION CONTENT	Health Information		X		X				
	Extra Context					X			
	Minimal Context	X		X			X	X	X

The variables that define the different survey versions (as listed in Table 1) are explained and described below:

Response Frame

WTP All the survey versions except for version H contained a willingness to pay (WTP) question. The WTP question asked respondents if they would pay for a one-step improvement in air quality, (whether this improvement was for air quality, health, or both depended on survey version). If they were not willing to pay, they were asked why; and if they were willing to pay, they were asked for a maximum dollar amount. For all of the WTP questions, this was a yearly household payment.

WTA Respondents to versions A, B, and E were also asked a willingness to accept compensation (WTA) question. The WTA question was identical to the WTP question, except that the WTA version asked respondents if they would be willing to accept compensation for a decrease in air quality, and the least they would be willing to accept monetarily to allow such a decrease.

Health vs. Visibility

Three In surveys C and D respondents gave WTP values for health, visibility, and total air quality. Respondents to the other survey versions gave values for air quality as a whole.

% Split In versions A and B, after responding to the WTP question, respondents also indicated the percent of their total air quality WTP value that they would attribute to health effects and the percent that they would attribute to visual air quality.

Form of Value Questions

Standard CVM Versions A, B, C, and D have WTP/WTA questions that follow the standard CVM procedure, with no added context.

Voting In accordance with the recommendations of Mitchell and Carson (1989) and of Fischhoff and Furby (1989), version E included added context for the WTP/WTA questions, containing more specific information about the connection between the WTP/WTA values and improvements/reductions in air quality. Instead of simply asking for payment/compensation for changes in air quality, the version E scenarios detailed two referenda (Mitchell and Carson, 1989) that would affect air quality as well as taxes and/or prices for which respondents could vote. Respondents in the WTP question are asked if they would consider voting for a referendum. If they answer “yes” they are asked what is the most they would pay before changing their mind.

Choice In version H, respondents did not answer a WTP question. Instead, they indicated whether they would choose an upgrade in air quality over an upgrade in either of two commodities: a television set and a camera. Photos of two levels of air quality and two levels of quality of each commodity were shown in the survey booklet.

Description of change in air quality

Average air quality change All of the versions except for version F asked respondents to value a one-step average increase/decrease in air quality.

Frequency distribution of air quality change Version F asked respondents to give WTP values for two types of increases in air quality, as expressed in frequency distributions of the number of days in a winter at given air quality levels. One frequency distribution showed an increase in very good days, with no change in bad days; the other frequency distribution showed a decrease in very bad days with no change in the good days. Each shift in the frequency distribution corresponded to a one-half step increase in the air quality ladder.

Context/Information Content

Health information Versions B and D contained information on the probable chemical make-up (carbon monoxide, sulfur dioxide, and ozone) of the air depicted in the eight pictures (the six rated plus two example pictures) on the color insert. The information was scaled in accordance with the U.S. Environmental Protection Agency's Office of Air Quality Planning and Standards' (OAQPS) "Air Pollution Index", so that respondents would know the extent each day violated air quality standards on the three chemical variables. Also, versions B and D provided descriptions of the negative health effects of violations of these standards.

Extra Context Version E contained more contextual information surrounding the valuing questions than did the other versions. Before answering the WTP/WTA questions, respondents read information about the Better Air Campaign and the ways in which it could help make the air quality better in Denver. Then, as explained under "Voting" above, respondents answered the WTP/WTA questions in the context of referenda that would result in better/worse air quality and that would cost/pay a certain amount in taxes and prices.

Minimal Context Versions A, C, F, G, and H, presented respondents with little information about the health effects of the air, the manner in which air quality can be improved, and the ways in which their money could change air quality.

Three Commodity Comparison Survey versions G and H, unlike the other survey versions, asked respondents to value two commodities in addition to air quality. This version asked respondents to consider a trade up of one step in air quality (which, except for the use of the word "trade", is the same valuation in all the air quality WTP questions across versions), as well as a trade up from a less expensive to a more expensive camera and television set. The survey booklets contained photos depicting the less desirable and more desirable air qualities, TV's, and cameras. The particular commodities and photos were chosen using pilot data, so that, on average across respondents, the camera trade would be slightly lower and the TV trade would be slightly higher in value than the air quality trade. Versions G and H differed on how the commodity and air quality trades were evaluated. In version G, respondents answered the usual form of WTP question for each of the three trades (camera, TV, and air quality). In version H, respondents reported choices for each of the two commodities that indicated which they would pay more for:

1. the air quality trade,
2. the same, or
3. the camera/TV trade.

Survey Mailing

Surveys were designed consistent with the Dillman Total Design Method (Dillman, 1988). The Dillman recommendations involve personalized mailing including a hand-signed cover letter, hand-stamped envelopes, follow-up reminder postcards, and a second mailing to households that did not respond to the survey following the first mailing. Also, the surveys were printed and folded into a booklet measuring 8 inches by 6 inches. The surveys were six (versions A and C), seven (versions B and F), eight(versions D and E) or 9 (versions G and H) pages long, including the cover and space for comments.

Each household in the sample received a version of the survey, a color photo insert, a cover letter, and an addressed, stamped envelope to return the survey to us. For the first mailing only, the package also included a two dollar bill, to thank respondents for their time and to encourage them to fill out and return the survey. In past survey research we have found that monetary incentives increase response rates significantly (Doane, et al., 1989).

The survey sample consisted of 1600 households (200 for each of the eight survey variations) randomly selected from the Denver Metro area, including the city of Denver and its neighboring suburbs of Arvada, Aurora, Littleton, and Wheat Ridge.

Response Rates

When corrected for bad addresses, the response rates ranged from 68.3% (version E) to 72.6% (version A). There was no statistically reliable difference in response rate by version (χ^2 (6) = .42, ns.). The overall response rate was 71%, which, when corrected for the 279 bad addresses, meant that 938 surveys were returned and used in the analyses.

1.4. Results

Variables

WTP The WTP values used for comparison across versions are either the value given for the WTP for a 1-step improvement in air quality (versions A, B, E, and G), the value given for the "total" WTP in the surveys with component and total WTP value questions (versions C and

D), or the sum of the two frequency distribution questions in version F. The frequency distribution questions each asked for a one-half step increase in air quality, so the sum of these two produced a WTP value comparable to the WTP values from the other versions. Version H did not include a WTP question, and so is not included in the WTP analyses.

Three types of independent variables were used in the regressions: (a) context design variables, (b) sociodemographic variables, and (c) air quality rating variables.

Context Design Variables Dummy codes were used to indicate the design features of each survey version. The variable HEALTH denotes whether or not the survey version contained the additional health information supplied by OAQPS; Versions B and D contained this information. The value of 1 for variable THREE indicates that the survey version contained three separate WTP questions: one for visibility, one for health effects, and one for total air quality; Versions C and D each have the three questions. The variable VOTING denotes Version E which framed the WTP question in terms of a referendum. The variable FREQ indicates Version F which displayed the improvement in air quality in terms of changes in frequency distributions. Finally, the variable COMMOD indicates Version G which asked the WTP question for air quality after asking similar questions for two other commodities–a camera and a television.

Sociodemographics Standard sociodemographic variables included AGE, GENDER (1 = male, 2 = female), EDUCation (see facsimile survey for categories), and INCOME. In past research on responses to risks, an important variable is whether or not there are children living in the household. The variable KIDS indicates whether or not there are any children 18 years or younger living in the household.

Air Quality Ratings Several questions asked respondents to evaluate the typical air quality they experienced and how much the visibility and health problems bothered them. VISRT is the rating on a seven-point scale of “how bothered have you been by the Brown Cloud’s effect on what you can see in the distance (mountains, buildings, etc.)?” HEALTHRT is the rating on a seven-point scale of “have you or your family been bothered by any health problems which you believe to be caused or aggravated by Denver’s air pollution?” In both cases higher ratings mean the respondent was more bothered. Finally, AQRATING is the respondent’s rating of the “current overall air quality (thinking of both health and visibility) on an average winter day” on the air quality ladder defined by the photographs enclosed with the survey.

Bid Refusals

The overall return rate for both mailings of the survey (including the postcard reminder) was 71% after adjusting for bad addresses. Although some respondents were asked for WTA as well as WTP values, only WTP was used for statistical analyses, because so few respondents gave finite WTA responses. Out of the 342 respondents who returned versions of the survey with the

WTA question, the vast majority indicated that they would not accept any amount of money for a decrease in air quality (i.e., their WTA value was "infinite"). Only 28 (8.2%) respondents gave finite values for WTA.

Out of the 812 survey responses to versions A-G (the versions requesting WTP values), 283 (35%) of those respondents did not give positive WTP values. Three (1.06%) of these respondents indicated that they would not give a positive WTP value because air quality had no value for them. The rest of the respondents either stated that they were not responsible for the problem or gave no reason. If more respondents had given zero dollar WTP bids for air quality, with the indication that the bid represented an actual zero value, then it would have been appropriate to perform a Tobit analysis to include these zero dollar bids in our regressions.

It is reasonable to conclude that, for many of the respondents, the refusal to bid indicated some other sentiment besides an actual zero value for air quality. Further, it is of interest to be able to predict when respondents are likely to give a positive WTP value, as well as to be able to include this 34% of the sample in the calculation of predicted values. A Probit analysis was performed to accomplish two goals: to identify variables that predict whether respondents will bid a positive WTP amount, and to provide a selection bias variable that would allow the inclusion of missing respondents in the model to predict WTP values. The three respondents who gave "actual" zero dollar values for WTP were not included in the Probit analysis.

Probit Results

Table 2 shows the results of the Probit analysis. AGE, INCOME, EDUC and KIDS all identified respondents who were likely to give a positive WTP value. Old age, low income, low education level, and having many children all decrease the chances that a respondent will give a positive WTP value for air quality. Note that neither concern for visibility VISRT or health HEALTHRT significantly increased the probability of a positive WTP. These variables would be expected to have a positive effect if refusal bids were indicative of zero WTP values. Rather, this result suggests that these respondents do not accept responsibility for the problem. Note that if either of these concern variables is included in the positive WTP regression reported below they show a positive and significant correlation with WTP.

WTP Distribution Figure 2 shows the distribution of refusal (zero) and WTP values, pooled across versions. The WTP value distribution is positively skewed, with a very long and thick positive tail, meaning that some respondents indicated that they were willing to pay quite large amounts for an improvement in air quality. When untransformed WTP values are regressed on the context and demographic variables, (see Table 3 for results of all WTP regressions), the resulting error distribution is also quite positively skewed (Figure 3, first distribution).

Table 2

Probit Analysis

(Parameters in bold are significant at p<.05)

Variable	Parameter Estimate	t
INTERCEPT	.214	.494
THREE	-.072	-.665
VOTING	.130	1.241
HEALTH	-.000	-.001
COMMOD	.244	1.632
FREQ	.017	.118
VISRT	-.001	-.011
HEALTHRT	.037	.894
AVERT	-.041	-1.349
AGE	**-.010**	**-2.471**
GENDER	.085	.825
INCOME	**.070**	**2.095**
EDUC	**.049**	**3.031**
KIDS	**-.17**	**3.268**

The WTP distribution appears to have a log normal shape, indicating that a log transformation of the WTP data would result in a more normal error distribution. Indeed, the regression of log(WTP) on the independent variables results in a more normal error distribution (Figure 3, second distribution), although this distribution exhibits a slight negative skew. In order to make this log(WTP) error distribution normal, some of the lower WTP values would have to be trimmed. A more satisfying solution is to transform the data using a Box-Cox procedure, which transforms the data using the best fitting α, such that Box-Cox = $(B^{\alpha}-1)/\alpha$.

When the Box-Cox WTP values are regressed on the independent variables, the resulting error distribution is symmetrical (Figure 3, bottom distribution). The estimated a is .2135. An a of 1 produces a linear transformation of the data; an a of 0 results in logarithmic transformation.

Regression Results

After the Probit analysis, three regressions were run, all of which included the selection bias variable (λ) from the Probit analysis. The first regression used untransformed WTP values, the second used log transformed WTP values, and the third used the Box-Cox transformation. The mean predicted values from the three regressions are presented in Table 4 while each of the three regressions are presented in Table 3. As the table of means shows, the mean for the model using untransformed WTP is higher than the means using transformed WTP. The mean for the log(WTP) model is somewhat lower than the mean for the Box-Cox model. Predicted values were computed for those respondents not giving a positive WTP value (the “nonbidders”); for the untransformed WTP regression, this mean predicted value is higher for the nonbidders than for bidders. For the log(WTP) and Box-Cox regressions, the predicted means are lower for nonbidders than for bidders.

In all of the regressions, the selection bias, l, variable is not a significant predictor of WTP values. Thus, in this case excluding bid refusals from the data analysis would not bias the estimated WTP equation.

Of the survey design variables, FREQ is the only reliable predictor of WTP values (at the 5% level), with COMMOD and THREE only weakly significant (at the 10% level) in the Box-Cox regression. Version F WTP values, used in the analysis of FREQ, were computed by adding two independent half-step values that respondents gave for two frequency distribution shifts. This effect could result from diminishing marginal valuation of air quality improvements, but more likely is the result of superadditivity.

Note that the upward bias in version C and D associated with THREE, where respondents first provided separate health and visibility values before giving a total WTP also indicates a super-additivity problem. Version G, in which respondents valued two commodity trades before valuing air quality, also resulted in marginally higher air quality WTP values. Respondents here may have anchored on the value of the commodity trades, feeling “morally” obliged to put more value on air quality.

INCOME and EDUCATION are the demographic variables that significantly predict WTP values, with higher income and higher educational level predicting higher WTP values (except for EDUC in the log(WTP) regression, which is below significance at the .05 level). Respondents’ ratings of a typical winter day (AQRATING) are predictive of WTP values in the linear regression only, suggesting that this rating is in fact not a reliable predictor of WTP values.

Commodity/Air Quality Comparisons In versions G and H, respondents were asked to evaluate trades in air quality, television sets, and cameras. Respondents to version G were asked to give WTP values for the three trades. Their mean WTP values for the camera trade were lower, overall, than their mean WTP values for the TV trade, and air quality WTP values were higher, overall, than either of the commodity WTP values.

To compare responses on the two survey versions, it is necessary to convert the Version G monetary values into imputed choices that are comparable to the Version H choice responses. The imputed choices are obtained by comparing the WTP values for the commodity trades and the air quality trades. If the respondents gave a higher value for the air quality trade than for the

Figure 3

Residuals for WTP Regressions

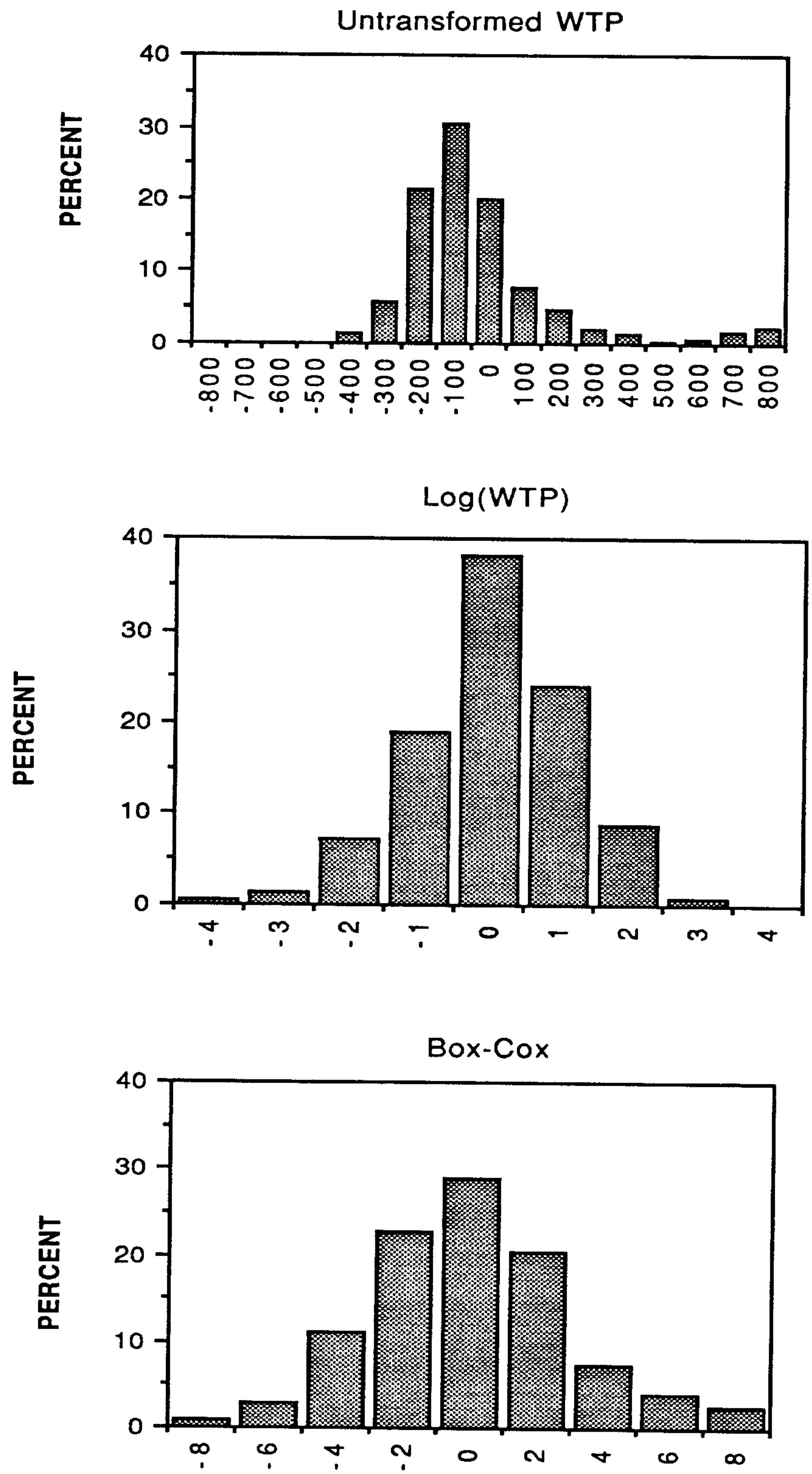


Table 3

Parameter Estimates and t's by Variable for Models of WTP

(Bold entries are significant at p < .05)

Variable	Untransformed	Log	Box-Cox
INTERCEPT	-134.374	**3.686**	**4.732**
	(-1.061)	**(5.399)**	**(3.114)**
HEALTH	-21.092	-.226	-.410
	(-.630)	(-1.254)	(-1.021)
THREE	**61.349**	.268	.651
	(1.807)	(1.465)	(1.599)
VOTING	34.302	-.115	-.239
	(.788)	(-.492)	(-.458)
FREQ	**84.691**	**.421**	**1.160**
	(1.991)	**(1.838)**	**(2.273)**
COMMOD	57.748	.216	.798
	(1.424)	(.990)	(1.641)
AGE	-.210	-.003	-.001
	(-.149)	(-.464)	(-.079)
GENDER	2.854	.015	.117
	(.101)	(.100)	(.348)
EDUC	**20.135**	.067	**.211**
	(1.997)	(1.235)	**(1.743)**
INCOME	**23.311**	**.104**	**.286**
	(4.077)	**(3.373)**	**(4.179)**
KIDS	-19.191	-.047	-.069
	(-.744)	(-.342)	(-.223)
AVERT	**-36.104**	-.119	-.333
	(-2.052)	(-1.258)	(-1.580)
LAMBDA	94.871	-.423	-.461
	(.710)	(-.588)	(-.287)
	$R^2 = .09$	$R^2 = .11$	$R^2 = .14$

Table 4

Mean Model Estimates of WTP

(Nonbidders did not give a positive WTP bid)

	Untransformed	Log	Box-Cox
<u>Separate Means</u>			
Nonbidders	244.22	52.68	82.77
Bidders	219.29	94.71	119.72
<u>Grand Mean</u>			
Nonbidders' bids counted as $0	146.30	63.18	79.87
Using Nonbidders' Pred Values	231.76	73.70	101.24

commodity trade, they are given a "1" (prefers air quality trade to camera/TV trade), if they gave the same value for the air quality and commodity trade, they are given a "2" (value the trades equally), and if they gave a higher WTP value for the commodity trade than for the air quality trade they are given a "3" (prefers commodity trade to air quality trade) for that trade.

Figure 4 shows the percentage of respondents in the choice categories, for the camera/air quality trade comparison and Figure 5 shows the comparable percentages for the TV/air quality trade comparison. Although most respondents favor the air quality trade over the commodity trades, there is a significant difference in choices due to version type ($\chi^2(1) = 14.3$, $p < .0001$ for the camera/air quality comparison and $\chi^2(1) = 23$, $p < .0001$ for the TV/air quality comparison). The choice differences due to reward type indicate a higher relative valuation of the commodity trades, and a lower relative valuation of the air quality trade, when respondents receive Version G of the questionnaire. Specifically, respondents in Version G are relatively more likely to value the camera/TV trade more than the air quality trade, and respondents in version H are more likely to value the air quality trade above the TV trade.

When the analyses are run using just ratings 1 and 3 (i.e., no ties), the version/rating difference is significant for the camera ratings ($\chi^2(1) = 13.22$, $p < .001$) but not for the TV ratings.

Figure 4

Camera Versus Air Quality Ratings by Version

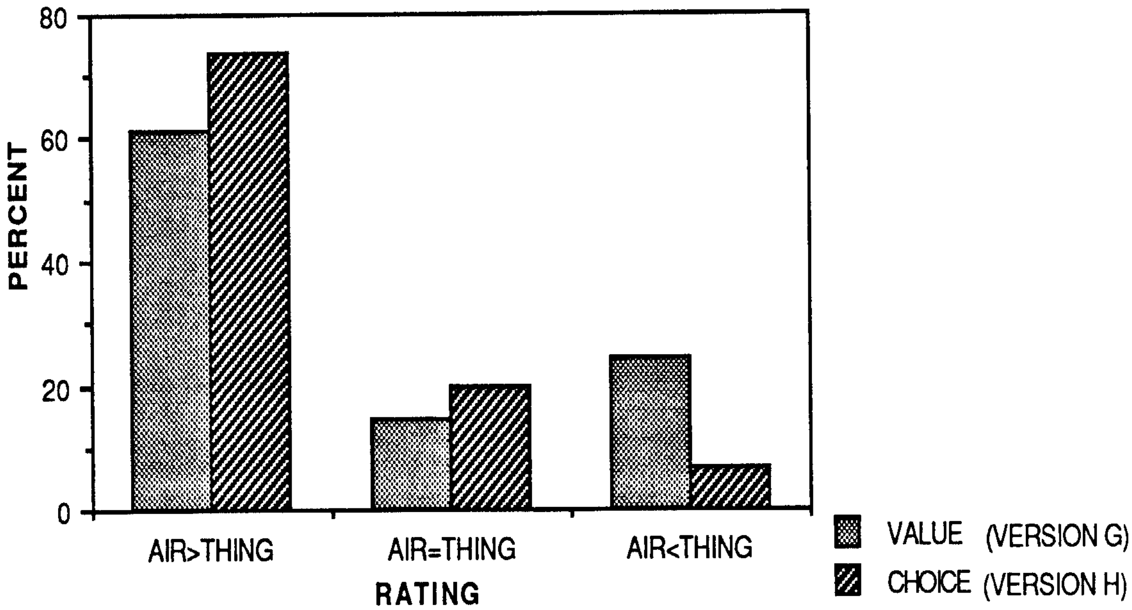
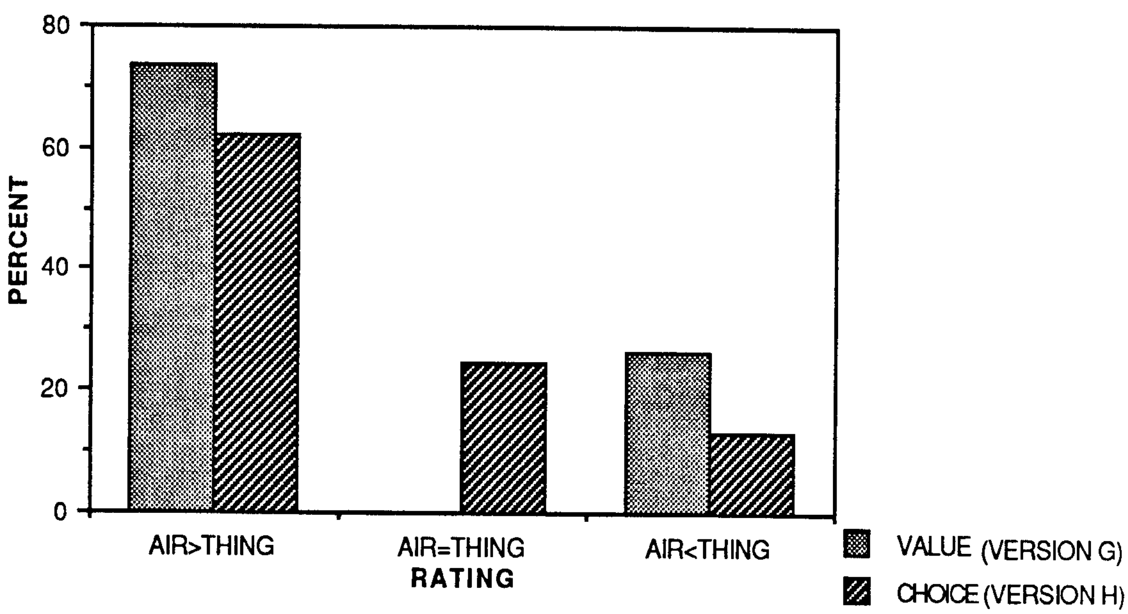


Figure 5

TV Versus Air Quality Ratings by Version



Ratings

For the "how bothered are you..." questions (VISRT and HEALTHRT), the average rating on a seven-point scale for visibility, 5.4, is reliably greater than the average rating for health, 3.3, ($t(377) = 30.9$, $p < .0001$). The correlation between the two ratings is statistically reliable ($r = .37$, $p < .0001$).

All respondents also rated six photographs on an air quality ladder separately in terms of visibility and health. For all the photos, the difference between ratings for visibility and health are generally small, with mean differences ranging from -.08 to -.19. The overall visibility-health rating differences is nevertheless statistically significant ($t(827) = 7.68$, $p < .0001$).

WTP Values within Survey Version Several of the survey versions asked respondents who gave WTP values to give the percentage of that value attributed to visibility and health. The mean share for visibility (27.2%) was reliably lower than the mean share for health (48.3%, $t(229) = -8.53$, $p < .0001$). 92% of respondents who gave positive numbers for both values gave percentages that added up to 100; the majority of the remaining respondents gave percentages that added up to less than 100%. Twenty-three percent of respondents giving values for both percentages gave values of "0" for both. Thus, respondents attributed 27.2% of their total value to visibility, 48.3% to health, and 24.5% to other sources of value. If we use the Box-Cox predicted values, including those for non-bidders, then the mean total bid of $101.24 would be distributed as follows: $27.54 to visibility, $48.90 to health, and $24.80 to other. Note that this approach implies a significant difference between health and visibility values and avoids the superadditivity problem discussed below.

Two of the survey versions asked respondents to state separate WTP values for visibility and health for a one-step improvement on the air quality ladder. Although the mean for separate visibility values was lower than for separate health values, the difference was not statistically reliable ($t(191) = .78$). The sum of respondents' visibility and health WTP values significantly exceeded their total WTP values for air quality ($t(190) = 2.51$, $p < .01$). Thus the superadditivity hypothesis is strongly supported in our results. Note, however, that the total WTP as given by respondents was only marginally significant and greater than in other survey versions.

The final WTP analysis compared respondents' WTP values for the two frequency distributions presented in version F. One frequency distribution depicted a half-step improvement in air quality involving a reduction in the number of bad air quality days; the other distribution depicted a half-step improvement in the number of good air quality days. The WTP values for the reduction in bad days were significantly higher than the values for increasing the number of good days ($t(104) = 2.08$, $p < .04$).

1.5. Conclusions

The first goal of this study was to attempt to apply an understanding of hypothetical bias arising from lack of market experience obtained from laboratory experiments to a field application of the CVM. Mean estimated values using the Box-Cox transform which attempts to eliminate hypothetical bias appear to be substantially smaller than those obtained by using a linear regression approach. In the case of air quality in Denver, selection bias resulting from bid refusals was not important. The treatment of refusal bids, assuming refusal bids are zeros or using the predicted values for non-bidders, was of substantially lesser importance than correcting for the skewed error distribution.

A second and related goal of this study was to provide information about the effect of contextual variables on values elicited by CVM questions about air quality. Our overall conclusion is that, contrary to the expectations of theorists such as Fischhoff and Furby (1989), for the most part, contextual variables in our survey did not greatly affect respondents' WTP values, implying that values are crystallized in this particular case. None of the observed context effects changed the order of magnitude of air quality values.

Respondents to the version with air quality frequency distributions did give significantly higher WTP values. In this case respondents evaluated two frequency distributions, each representing a half-step change in air quality, so that the sum of their two values would theoretically equal a WTP value for a one-step increase in air quality**.** However, since this was an attempt to split out the value of reducing the number of poor days from the value of increasing the number of good days, it is not surprising that the sum of the values exceeded that obtained in other versions given the problem of superadditivity. Note also that the sum of separately obtained health and visibility values exceeded the total value obtained from the *same* respondents. Thus superadditivity appears to be a consistent problem for respondents that requires great care in the design and interpretation of surveys attempting to obtain values. Use of percent splits rather than use of separate values to obtain component values may be an appropriate design strategy.

The remainder of this report is organized as follows: Chapter 2 discusses issues in contingent valuation; Chapter 3 outlines the design of the field experiment in contingent valuation methods; Chapter 4 provides descriptive statistics, and Chapter 5 is an analysis of our field data with respect to hypothetical bias and context effects.

CHAPTER 2
Issues in Contingent Valuation

2.1 Introduction

Two primary issues remain in the application of the CVM. First, the nature and extent of hypothetical bias has proven difficult to quantify. Survey values obtained in the field tend to be bimodally distributed with a large number of zero bids and an upper mode which is skewed, showing a thick tail of very large bids. Researchers have tended to view both the large number of zero and very large bids with considerable skepticism. Fortunately, laboratory experiments can shed considerable light on this problem. Second, if context effects are important, i.e., if different survey designs obtain very different values for the same commodity, then estimated values are not robust. The two issues are obviously related in that if context has a large impact on values, then hypothetical bias is likely severe – especially in cases where seemingly trivial or irrelevant changes in context have a large impact on values. We consider each of these issues in turn and conclude this section by suggesting a specific set of econometric procedures.

2.2 Hypothetical Bias

A consistent and striking pattern has been seen in the results of laboratory experiments dealing with hypothetical values. In the laboratory, hypothetical bids obtained from subjects for a commodity showed an increased variance relative to actual bids. Further, increasing market experience (repeated rounds for a particular auction institution) and increasing incentives (increased payoffs for participation in a particular market institution) there was a reduction in the variance of the bids.

The first experiment to compare hypothetical bids to auction behavior was undertaken by Coursey, Hovis and Schulze (1987). These experiments produced mean hypothetical bids which were statistically identical to mean experienced auction bids. The commodity used in the experiment was a bitter tasting liquid, sucrose octa acetate, which was unfamiliar to subjects. A careful description of this commodity was given to subjects first, and then they were asked how much they would be willing to pay (WTP) to avoid a taste experience. After this, subjects were allowed to taste the liquid prior to being asked for their WTP. In this second stage subjects were familiar with the commodity but had no real market experience with it. Finally subjects participated in a competitive auction.

Mean bids and their associated variance respectively, were as follows: Hypothetical with no experience \$2.60 and \$15.80; with experience with the commodity \$2.27 and \$5.06; and actual auction bids with market experience \$1.95 and \$5.23. Thus, variance was greatest for the totally hypothetical bid as opposed to the experienced auction bids. However, it appears that the decrease in variance was associated with experience with the commodity rather than with experience with the market institution.

This phenomenon can be seen clearly in Figure 2.1, which graphs the frequency distribution of bids in the Coursey, Hovis and Schulze experiments. The top graph shows the number of bids made at various levels (categories). These bids represent those made for the completely hypothetical case. There are 9 bids at the zero level and a number which trail out to the right hand side of the distribution. The second graph shows the distribution of bids for the case where subjects had some experience with the good (Semi-Hypothetical). As discussed above, the number of bids had been reduced. Further, the tail to the right has also been pulled in thereby reducing the variance around the mean of the bids.

The last graph shows the distribution of bids for the actual competitive auction (End Trial) distribution of bids. It is important to see, as noted above that there is little difference between this distribution and the one above it. The number of zero bids has decreased again, but the right hand tail is still the same. Therefore, we get an indication that experience with the good may have as much impact as actual market experience in lowering the variance of bids (seen as a reduction in the number of zero bids and the skew of the distribution).

These findings are verified by the results of Irwin, McClelland and Schulze. Subjects in this experiment were first given an initial cash balance. They then faced a 1% chance of a $40 loss. The loss was determined by the drawing of a chip from a bag which contained 99 white chips and 1 red chip. If the red chip was drawn each subject lost $40. Before the drawing, four insurance policies were sold. Each subject submitted his or her bid for the available policies, and the recipients were determined by a Vickery auction. Each subject was informed whether or not they received insurance and the draw was made.

Two treatments of this design were run: One real, which involved real cash balances and real cash losses; and the other hypothetical which involved hypothetical balances and losses (subjects received only class credit - no money). Within each of these treatments there were two variations. One involved 50 draws from the bag, and the other 150 draws.

The 50 round hypothetical treatment had a mean bid of $7.52 with a variance of $948.52. When the same number of rounds were used, but the bids were real the mean was $3.88, and the associated variance was $192.52. When the number of rounds was increased to 150 rounds the hypothetical and real means were $7.50 and $3.12 respectively. There associated variances were respectively $608.87, and $30.33.

Figure 2.1

Frequency Distributions

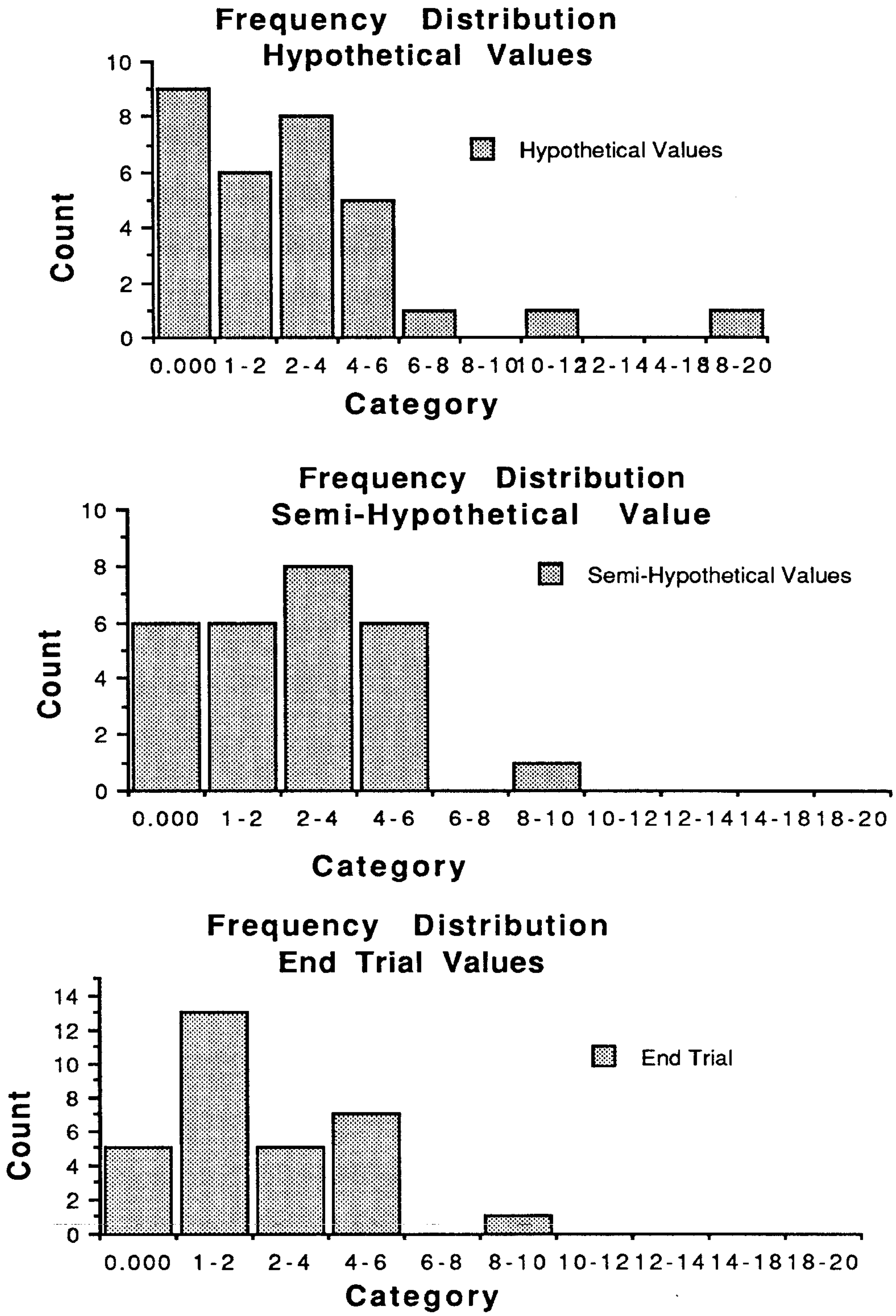


Figure 2.2 also shows the results of both the hypothetical and real treatments for the 50 and 150 round variations. The bid-expected value (BID/EV) ratio is presented on the horizontal axis. A log scale is used because many of the ratios are very large and because bids tend to be distributed log-normally.

As can be seen in the figure, the hypothetical bids, in both the 50 and 150 round variations, have a large distribution. As with the distributions shown in Figure 2.1, these hypothetical values have a large concentration at zero and are skewed out to the right. In this case the right hand tail is much more pronounced than in the previous case, which will have a more profound effect on the dispersion around the mean.

Looking at the results from the rounds where the real money was used, we see that the variance has decreased. As with the distribution of the experience bids above, these real rounds have many fewer 0 bids. Further, the right hand tail is much smaller. Therefore, it appears that giving subjects some real experience with a commodity decreases the variation in the values they form on that commodity.

Figure 2.3 shows how increasing variance in hypothetical bidding can bias estimates of actual behavior. First, in the top panel, if we assume that the negative values of the left tail of the hypothetical bid distribution produce zero bids (which are more common in hypothetical bidding than in actual bidding in auction experiments), then the mean bid of the hypothetical bid distribution will overestimate the mean of the actual bid distribution. This situation explains much of the decrease in mean bid in the Coursey, Hovis, and Schulze experiment. Second, in the middle panel of Figure 2.3 a skewed hypothetical distribution, relative to the actual bidding distribution is shown. Here the extended right hand tail is the source of an upward bias in the mean hypothetical bid. This source of error dominates the results of the experiment conducted by Irwin, McClelland, and Schulze (1989). The bottom panel shows a situation where the increase in hypothetical bidding variance produces an unbiased estimate of mean bid as occurred in the experiments conducted by Cox, Smith and Walker (1989) which varied experimental inducements.

Given the experimental evidence summarized above, what model can be used to explain hypothetical bias which might occur in field surveys? Assume for simplicity that all individuals have the same true willingness to pay, W. However, the bid they provide in response to a hypothetical question about willingness to pay is B. Then

$$B = W + \epsilon \quad (\text{for } W + \epsilon \geq 0)$$

where ϵ is hypothetical error with a frequency distribution $f(\epsilon)$. The appropriate question in this

Figure 2.2

Experimental Values

Source: Irwin, McClelland, Schulze (in press)

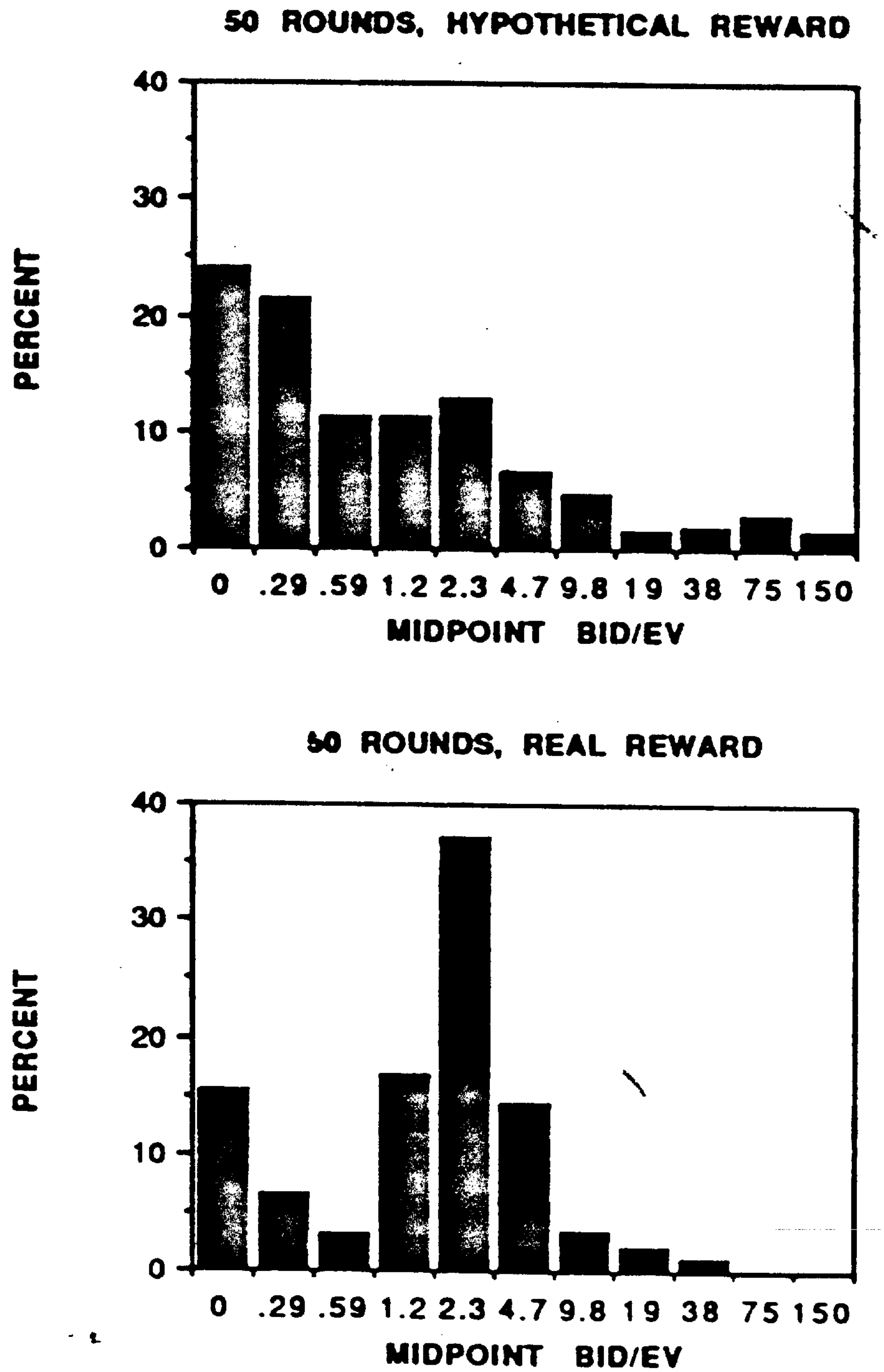
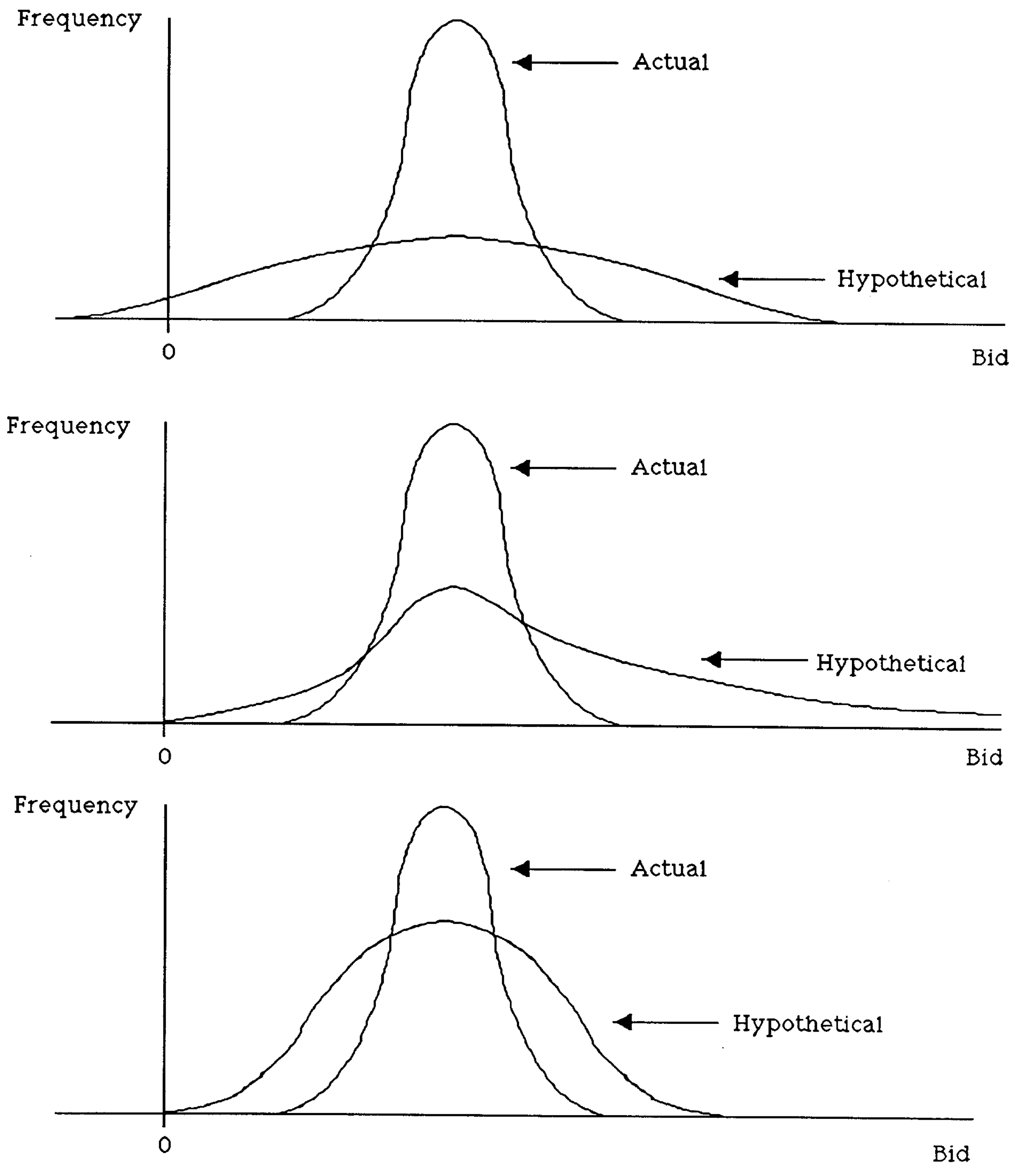


Figure 2.3


Frequency
Actual
Hypothetical
0
Bid
Frequency
Actual
Hypothetical
0
Bid
Frequency
Actual
Hypothetical
0
Bid

simple situation is, when will the mean of B,

$$\bar{B} = \int_0^\infty f(\varepsilon)\, B\, d\varepsilon,$$

be an unbiased estimate of W — or, in a more realistic situation, when will the mean of predicted values from a regression explaining WTP (using income, etc.), produce an unbiased estimate of mean WTP. As noted above, two sources of error can arise.

First, even if we assume that hypothetical error is symmetrically distributed so

$$\int_{-\infty}^{+\infty} f(\varepsilon)\, \varepsilon d\varepsilon = 0, \qquad (1)$$

a significant error can arise because of truncation bias. Since negative values of $W + \epsilon$ imply corner solutions in the consumer choice problem ($B = 0$ for $W + \epsilon < 0$), zero bids result from large negative errors. Hypothetical error in estimating W will thus be positive and equal to

$$\bar{B} - W = \int_0^\infty f(\varepsilon)\, \varepsilon d\varepsilon > 0$$

and will increase with an increase in the variance of $f(\epsilon)$. Thus, use of the mean bid to estimates WTP will overstate the true value. Similarly regression estimates used to predict WTP will be upwardly biased unless Tobit is used to account for truncation (Madala, 1983).

Second, $f(\epsilon)$ may not be symmetrically distributed. As noted above, available data both from the field and laboratory suggest that $f(\epsilon)$ is often skewed. Assume for purposes of illustration that e is distributed log normally so

$$\ln B = \ln W + \epsilon.$$

The simple mean $\bar{B}$, will be upward biased, but the mean of lnB will be an unbiased estimator of lnW since (1) still holds. This implies that a transformation of survey bids (such as the natural logarithm) which produces a symmetrical error distribution may well reduce hypothetical bias by eliminating skew. The Box-Cox (Box and Cox, 1964) transformation, $(B^{\propto} - 1)/\propto$ has proven extremely useful where $\propto$ is selected by the procedure to normalize the error distribution in regression analysis. Further, it incorporates both the linear ($\propto = 1$) and natural logarithm ($\propto = 0$) transformations as possibilities. Use of this procedure has several advantages. First, in the past large suspect bids obtained in the CVM have been removed through trimming (e.g., Desvousges, Smith and Fisher, 1987). Trimming procedures remove large outliers which deviate from an estimated linear regression model by exceeding some predetermined statistical threshold. However, in the situation we have described the mean of predicted bids falls as that threshold is lowered, making final estimated values dependent on the threshold chosen. If skew is present, the procedure we propose will also lower mean values if bids predicted by the estimated regression equation are used in calculating the mean. However, the reduction in predicted mean bid will be determined by the optimal choice of $\propto$, the Box-Cox parameter, so as to set the mean of the residual error in the predicting equation equal to zero. If hypothetical error dominates the residual, then it is obviously desirable to use an estimating procedure which does not bias the estimated coefficients through a skewed hypothetical error distribution.

A final problem in the interpretation and analysis of contingent values – which we feel is unresolved at this point in time – is the presence of protest zero bids or refusals to bid when respondents are asked for WTP. When pretesting survey instruments, researchers in debriefing respondents have often found that zero bids are not associated with a zero value to the respondent, but rather the respondent does not feel responsible for the problem. For example, a respondent may argue that although she is harmed by air pollution, she is not responsible for creating the problem (e.g., she does not own a car). Rather, industry and others should pay and are morally responsible for cleaning up the problem. Interestingly, such respondents when asked for their willingness to accept to allow a decrease in environmental quality often refuse any amount of money, arguing that to do so would be morally wrong. Thus, moral reasoning results in an unwillingness of respondents to provide any tradeoff between money and the public good in question. An L-shaped indifference curve between money on the vertical axis and the public good on the horizontal axis results from the application of moral reasoning, a situation similar to that described by Hahneman (1989). As argued by Smith and Desvousges, 1987, the absence of positive bids from such respondents results in a potential selection bias problem since as many as 30% of respondents may refuse to provide values. In estimating a regression model for those respondents who do provide a WTP, selection bias must be accounted for (Heckman, 1979) to obtain unbiased coefficients. However, a question remains as to whether or not to include predicted values from the resulting estimated equation for those individuals who refused to bid. We propose that mean bid should be calculated two ways, first assuming zero values for those who refuse to bid and second using predicted values for these respondents. Note that identification of bid refusals can be accomplished in at least three ways in the design of a survey instrument. First, a question asking why the respondent bid zero can be included. Second, a question asking for willingness to accept for a decrease in the level of provision can be included along with the WTP question. Third, questions asking how concerned about, bothered by, or important the commodity is to the respondent can be used to check for the consistency of a zero bid.

2.3 Context Effects

A potential problem in the design of any survey instrument is the degree to which the wording of the survey can affect respondents' answers. Both the wording of the valuing question and the information surrounding the question, which we term the *context* of the question, can affect the value given. Hogarth (1982) in an edited volume presents a number of papers that confirm the notion that context can have an affect on people's values, even in situations in which the context should logically have no effect. For example, researchers (e.g., Noell-Neumann, 1970) have found that the order in which (independent) questions are asked can affect people's answers to the questions. Other researchers (e.g., Tversky and Kahneman, 1986, Lichtenstein and Slovic, 1971, 1973) have found that how the question is expressed (e.g., in terms of losses versus gains, or in terms of percentages instead of fractions) can affect people's responses.

In order to understand context effects, it is helpful to think of values as being more or less "crystallized" (Schuman and Presser, 1981). If a person has had the opportunity to think about and/or declare a value for a commodity (in a marketplace, for example) to such an extent that the value is "set" in the mind, then it is unlikely that the manner in which the value is elicited will affect the value. In such a case, we would say that the commodity's value is "crystallized", and relatively impervious to context effects. If, on the other hand, the commodity is one for which the person does not have a set value, because the commodity is not often traded in a marketplace and the person has not thought of the commodity in monetary term, then the value for that commodity is less crystallized. In such a case, context difference could affect the value, in two general ways.

One way in which context can have an effect is in the process of evaluating the commodity. For example, when a respondent reads the words "air quality", many components of the concept "air quality" may come to mind. In a sense, the words "air quality" themselves may have different meanings for different people, and for many people, the meanings will be quite vague and unformed. Context can help "flesh out" the concept of air quality, or place emphasis on different aspects of the problem. Evidence has shown (Tversky and Kahneman, 1974) that whatever components of a commodity are most "available" to the mind of the respondent will figure most strongly in the evaluation of the commodity. For example, reading information about the health effects of particulates in the air in Denver could cause respondents to place more emphasis on the health dangers of air pollution, thus raising their values for improving air quality. It is as if the context provides an anchor, a way to think about the problem, for the respondent.

Another type of context effect is rooted in the difficulty that people have in assigning values to commodities such as “air quality.” Context can help respondents understand how the general concept of “air quality” can be translated to a monetary scale, especially since it is likely that many respondents have not thought of environmental commodities in monetary terms. It is important, for that reason, that CVM questions be given enough context so that respondents believe and understand that their money would actually buy the commodity they are evaluating. Some researchers (Fischhoff and Furby, 1989) have suggested that even seemingly minor wording differences in CVM questions could result in respondents valuing essentially different commodities, thus making interpretation of the results impossible. The disadvantage to giving too much contextual information of this kind is the danger that the respondents will have a response to the contextual information that is unrelated to their actual values for the commodity. For instance, asking respondents if they would be willing to be “taxed” to pay for “governmental programs” that would clean up the air pollution in Denver may result in angry refusals to give a value, not because the air pollution problem is unimportant to the respondents, but because of disgust for the government, or taxes, or some other element of the questions’ context.

Another explanation for the source of context effects can be illustrated by difficulties which have arisen in artificial intelligence (AI) research. Consider the problem of programming a computer so that it can be told to do something in conversational English and respond appropriately, as would a human being. The computer might well take everything that was said to it literally or confuse possible alternative meanings. To solve this problem AI researchers have turned to a cognitive model developed by Shank (for a non-technical description see Rose, 1985) which supposes that key words in a message invoke a mental script which incorporates or guides the appropriate response. Shank argues that few messages are likely to be sufficiently complete as to be understood or to allow proper action to be taken without reference to such a script. The script serves to fill in missing information in the message. Such scripts are, of course, based on memories and incorrect scripts may be employed by mistake. Rose in his book describing AI research (1985) includes the following anecdote which illustrates the role of scripts:

Wilensky...became a fan of “The George Burns and Gracie Allen Show” – partly from an AI point of view, I was told by Chris Riesbeck, Schank’s research associate. Wilensky came into Riesbeck’s office one day and said he didn’t understand how people who didn’t study AI could understand Burns and Allen, because so many of the things that Gracie got wrong were pronominal references. Given a choice of two different meanings of he, she’d come up with the wrong one.

"An example that's interesting for people who're writing parsing programs," Riesbeck said, "is the show where she's at home and there's all these flowers around the room, and Blanche comes in – the next-door neighbor – and says, 'Gracie, where'd you get all these lovely flowers?' And Gracie says, 'Well, Helen was in the hospital, and George said I should visit her and take her flowers – so when she wasn't looking, I did!'

"Now, that's the kind of mistake a program ought to make," Reisback continued. "But why is it that it's a joke? Clearly you don't think of that as the average interpretation. When you go to a hospital, you have a script that says 'bring flowers.' Since you know you're supposed to bring flowers, you parse 'take her flowers' directly as 'take flowers to her.' That's an AI way of looking at it."

Clearly, in a survey asking for contingent values, context effects can occur because different words, different information, even inclusion of irrelevant alternatives can all call up different scripts which will influence choices. Of course, the more clear and complete the message or question is, the more complete the context of the question will be and the less a respondent has to rely on scripts to fill in missing information. This suggests that one approach in situations where values are uncrystallized (so context problems are likely to occur) is to attempt to fill in the context as completely as possible in the survey design. This is the approach recommended by Fischhoff and Furby (1988) in their review of the CVM as applied to visibility valuation. Specifically, they suggest:

1. The definition of changes in air quality must be clearly stated. A survey should be explicit concerning the attributes of air quality; the cause of air pollution; the source, extent, and timing of the proposed change; the reference and target levels of air quality; and the certainty of provision of the change.
2. Fischhoff and Furby require that the payment procedure be explicit concerning the payments' attributes, the context (or payment vehicle), who is to offer payment, the reference and target levels of payment, the extent and timing of payment, and the certainty of payment.
3. Finally, Fischhoff and Furby are concerned with the social context of the transaction. It should be clear from the questionnaire who else is involved in the transaction, what decision rule will be used to determine whether or not the improvements will actually be made, and what other issues are involved in the transaction.

In summary, context effects are not likely to be a serious problem (i.e. cause order of magnitude changes in measured values) in the case of crystallized values. However, in cases where values are not crystallized, a much larger burden is placed on the survey design to provide a complete, realistic and clear context for the formation of values.

The implications of the discussion presented above are straightforward. The degree of crystallization of values obtained in a particular CV study can be estimated by examining the robustness of values with respect to changes in context. Several survey variants should be employed which vary the context in dramatic but not misleading ways. This approach has been applied by the Colorado group to the problem of valuing electric power outages for the Niagara Mohawk Power Corporation in New York State. Six survey variants were employed. The base treatment obtained values for one outage defined in terms of amount of advance notice, time of day of occurrence, length, season, etc. Other treatments included: (1) a variant with five outages and another variant with ten outages (to test for respondent fatigue and order effects); (2) a variant which first reviewed the impact and response to a past "bothersome" outage before asking for values and another variant which conducted the same review both for a past outage and for the hypothetical outage to be valued in the survey (to test the impact of additional thought on values); and (3) a variant measuring both WTP and WTA.

In order to test the degree to which values are crystallized, as well as the degree to which respondents need contextual information to make sense of CVM questions, we propose that context be varied across different versions of the survey. If values obtained across survey versions are robust with respect to changes in context, then it is likely that values are crystallized and hypothetical bias may not be a severe problem for the particular commodity.

Another specific difficulty associated with the way in which values are obtained concerns the hedonic nature of many public commodities. For example, air quality improvement provides value both from reduced health effects and from improved visibility. Fischhoff and Furby (1987) suggest that people may not be able to provide component values in such cases because they have only considered air quality as a whole. Mitchell and Carson (1989), based on pretesting with focus groups, argue that people treat health improvements and visibility improvements as inseparably related. If this is the case, even if the physical situation allows health and visibility to be improved separately by controlling different pollutants, respondents will provide values as if they are joint products. If attempts are then made to separately value one component, respondents will provide a value for both components since they assume that components of value move together. Values obtained separately will add up to more than that obtained by directly asking respondents for the total value of air quality improvement. Thus a superadditivity problem will exist. One way to overcome this difficulty is to ask respondents to split total value between health and visibility after they have given a total value. This approach is consistent with their "maintained view" that these are joint products. Unfortunately, however, superadditivity may be a more pervasive problem in that, if respondents have not previously had experience in breaking down values, stated components of value may not generally add up in a logically consistent way. It is important to test for logical consistency by varying context to check for superadditivity problems. If such a problem exists then the survey can be designed to assist respondents in deriving logically consistent answers (e.g., percent splits of a total value).

CHAPTER 3

Design of Field Experiment in Contingent Valuation Methods

3.1 Introduction

An important issue in the design and application of the CVM is accurately describing to people the commodity they are valuing. Many surveys implement a scale which represents different levels of provision of the commodity, but it is often difficult to ensure that subjects are able to understand and reliably use such scales. For example, it is possible that individuals con not distinguish between differences in photos that depict varying levels of air quality. In an effort to detect such measurement and perception problems, we attempted to provide all respondents with uniform perceptions of air quality. Respondents could then be asked to evaluate changes in air quality, as shown on the pictures provided, according to their own air quality scales. In particular, we asked all respondents to view a set of six color pictures of the Denver metro area and rate each of the pictures on a 7-point air quality ladder. We provided two example pictures that were previously rated to serve as anchors (the example pictures were rated 2 and 6, respectively). By rating the pictures themselves, respondents could form an idea of what a “step” on the air quality rating ladder consisted of. This enabled respondents to better answer questions later in the survey that referred to a one-step increase or decrease in air quality.

3.2 Survey Design

In an effort to test the effects of survey design on air quality values, eight different surveys were designed. Each embodied a combination of variables that were of interest. This section contains descriptions of the general design aspects being explored in each of the eight survey versions (see Figure 3.1 for a summary of the survey designs), first outlining the various elements of the surveys under the heading of survey variables, and then comparing and contrasting the different survey designs under the heading of survey variants. Figures 3.2 through 3.12 are located at the back of this chapter and contain reproductions of the survey questions referred to in this discussion.

Survey Variables

Response Frame

WTP All survey versions with the exception of version H contained a willingness to pay (WTP) question. Respondents were asked whether they would be willing to pay for a one-step improvement in air quality. If they indicated that they would be willing to pay, they were asked for a dollar amount expressed in dollars per year for their entire household. If they were not willing to pay, they were asked to provide a reason why.

Table 3.1

Summary of Brown Cloud Survey Design Features

		BASE		SUPER ADDITIVE		VOTING	FREQ. DIST.	3 COMMODITY COMPARISON	
	VERSION	V-A	V-B	V-C	V-D	V-E	V-F	V-G (WTP)	V-H (Choice)
GAIN/ LOSS	WTP	X	X	X	X	X	X	X	
	WTA	X	X			X			
HEALTH vs. VISIBILITY	Superadditive (3 quest.)			X	X				
	% Split	X	X						
FORM OF THE VALUE QUESTIONS	Std. CVM	X	X	X	X		X	X	
	Voting					X			
	Choice								X
DESCRIPTION OF CHANGE IN AIR QULAITY	Average Air Quality Change	X	X	X	X	X		X	
	Freq. Distribution of Air Qual. Change						X		
CONTEXT/ INFORMATION CONTENT	OAQPS Health Information		X		X				
	Extra Context					X			
	Minimal Context	X		X			X	X	X

WTA Respondents to versions A, B, and E were also asked a willingness to accept (WTA) compensation question. The WTA question was identical in format to the WTP question, except that respondents were asked if they would be willing to accept compensation for a decrease in air quality, as well as the minimum amount of compensation they would require.

Health vs. Visibility

Superadditive (3 quest.) Respondents to surveys C and D were asked to give separate WTP values for health, visibility, and total air quality (see Figure 3.4). Respondents to the other survey versions were asked to give values for air quality as a whole.

% Split Respondents to versions A and B were asked to divide into two parts the percent of their total air quality WTP value that they would attribute to health effects and the percent that they would attribute to visual air quality (see Figure 3.3).

Format Of Value Questions

Standard CVM Versions A, B, C, and D have WTP/WTA questions that follow the standard CVM procedure with no added context.

Voting In accordance with recommendations by Fischhoff and Furby (1989), respondents to version E were given added context for their WTP/WTA questions, containing more specific information about the connection between their WTP/WTA values and improvements or reductions in air quality. Instead of asking for payment or compensation for changes in air quality, the version E scenarios attempted to elicit values using a "referendum" question format. Respondents were asked how they would vote on a hypothetical city-wide referendum addressing trade-offs between air quality and taxes and/or commodity prices.

Choice In version H, respondents did not answer a WTP question. Instead, they indicated whether they would choose an upgrade in air quality over an upgrade in two commodities: a television set or a camera.

Description Of Change In Air Quality

Average air quality change When asked to provide a dollar value, all of the versions, except for version F, asked respondents to value a one-step increase or decrease in the average winter day's air quality.

Frequency distribution of air quality change Version F asked respondents to give WTP values for two types of half-step increases (as opposed to one-step) in air quality. Frequency distributions showing the average number of winter days at given air quality levels (e.g.,"good", "typical", or "poor") were used. One frequency distribution showed an increase in the number of "very good" days, with no change in the number of "bad" days; the other frequency distribution showed a decrease in the number of “very bad” days with no change in the number of “good” days. Note that the sum of the two values given would correspond to a one-step increase in air quality where the number of “very poor” days is reduced and the number of “very good” days is increased.

Context/Information Content

OAQPS health information Versions B and D contained information on the probable composition (carbon monoxide, sulfur dioxide, and ozone) of the air pollution depicted in the 8 pictures provided to the respondents. This information was scaled in accordance with the Office of Air Quality Planning and Standards and in accordance with the EPA’s established National Ambient Air Quality Standards published in the *Colorado Air Quality Data Report*, 1987, by the State of Colorado Department of Health, Air Pollution Control Division, Technical Services Program. The data and the informational text inserted in the surveys were provided by OAQPS (See Figure 3.3). The standards used were:

POLLUTANT	AVERAGING TIME	CONCENTRATION
Carbon Monoxide (CO)	8-Hours	9 ppm(10mg/m^3)
Ozone (0_3)	1-Hour	0.12 ppm (245mg/m^3)
Sulfur Dioxide (SO_2)	Annual Mean Arithmetic	0.03 ppm (80/mg/m^3)

This data enabled respondents to see to what extent each pollutant was present on the days that the photographs were taken. Versions B and D also provided descriptions of some negative health effects associated with violations of these standards.

Extra Context Version E contained more contextual information surrounding the value questions than did the other versions. Before answering the WTP/WTA questions, respondents were given information about the Better Air Campaign and the ways in which it could help improve Denver’s air quality. Then, respondents answered the WTP/WTA questions in the context of a referendum proposing better air quality but higher taxes and commodity prices, or, making air quality worse with lower taxes and commodity prices. (see “Voting” above)

Minimal Context In versions A, C, F, G, and H, respondents were given little information about air quality’s affects on health, the ways in which air quality can be improved, or the methods in which their money could change air quality.

Commodity Comparison Survey versions G and H ask respondents to value two commodities other than air quality. Respondents are asked to consider a trade up of one step in air quality after evaluating a trade up from a less expensive to a more expensive camera and television set. Respondents were shown photos depicting the less desirable and more desirable

T.V.'s, cameras, and air qualities. The particular commodities and photos were chosen using pilot data so that on average, across respondents, the camera trade should be valued slightly lower and the T.V. trade should be valued slightly higher in value than the air quality trade. A description of the pilot is provided later in this chapter under the section describing the survey variants G and H.

Versions G and H differed on how the commodity and air quality trades were evaluated. In version G, respondents answered the usual form of WTP questions for each of the three trades (camera, T.V., and air quality). In version H, respondents reported choices for each of the two commodities by indicating that they would pay more for: the air quality trade, the camera/TV trade, or that they were indifferent between the two trades.

Survey Variants

Common Aspects Across Versions

The "base" survey, version A, is reproduced in its entirety in Figure 3.2 and serves as the template on which all other versions were based. All surveys (V-A through V-F) have 4 sections: "The Issues", "How Do You Rate Denver's Air?", "The Value of Clean Air to You", and "About You". Introductory sections were similar across all variants. In an effort to focus respondents on the issue being addressed, all versions of the survey asked respondents to indicate how bothered they have been by Denver's brown cloud with respect to visibility and health impacts (V-A, Q-1 - Q-2). Next, they were asked to think about how often and where they have heard or read about the air pollution problem in Denver (Q-3). Following this set of questions, each survey had respondents look at a color photo insert. The insert contained six pictures of the Denver area during different visibility conditions as well as a seven-step visual air quality ladder to aid respondents in rating each picture. The insert also contained two example photos which served as anchors for steps 2 and 6 on the ladder. The photos were obtained from a unique, prior study of urban visibility perception (Stewart, et al., 1983, 1984). To ensure that the format used would not bias the presentation of various air quality levels, all photos were previously rated during a pre-testing exercise. Respondents were asked to rate each of the six photos using the seven step visual air quality ladder, first in terms of visibility, and then in terms of the healthiness of the air (Q-4 - Q-5). The aim of this exercise, as mentioned earlier, was to: 1) help define the commodity for the individual; 2) familiarize them with the seven point scale; and 3) induce thinking about the commodity to be valued later in the survey.

Another question common to all survey variants asked the respondent to think about the current overall air quality on an average winter day (the high pollution season) in the Denver metro area and to provide a rating for this average day (Q-6). This rating was used later in the value elicitation questions as a benchmark from which changes in air quality were defined to the respondent. Finally, the last section of each survey "About You" asked for demographic data (age, education level, etc.) about the household being surveyed (Q-11 - Q-20). This information was used to form a profile of the survey respondents. At the end of each survey, respondents were encouraged to use the given space to write their comments about the survey and Denver's air quality in general. The demographic questions are essentially the same across all versions of the survey, with the exception of Version E (figure 3.5) which contains two additional questions.

Unique Aspects Across Versions

Versions A and B These versions employ a minimalist design approach that is similar to earlier contingent valuation studies. However, in these two versions much more attention is focused on perception of air quality changes. At issue is the impact of additional information on respondents' air quality values, particularly the robustness of the values to changes in context. To test for these context effects, Version B contained additional health information (figure 3.3) provided by the Environmental Protection Agency's Office of Air Quality Planning and Standards (OAQPS) which was not included in Version A. The two versions were otherwise identical. As mentioned earlier, OAQPS provided not only the additional health information content, but the actual text and presentation format as well.

Versions A and B included *both* willingness to pay (WTP) and willingness to accept (WTA) questions (figure 3.2, Q-7 - Q-10). These questions were constructed in the manner of a standard CVM format in which an outline of the environment is provided within the text of the question. The respondent is then asked to make a decision within this hypothetical environment. These questions allowed for YES or NO responses. Respondents who answered "YES" were asked to provide the dollar amount they would be willing to pay (accept) each year for a one step increase (decrease) for both better (worse) visibility and fewer (more) health problems on an *average* winter day in comparison to the rating they had provided earlier for the average winter day. Finally, respondents were asked what percentage of their answer was for health and what percentage for visibility.

The CVM questions were specifically designed to allow respondents to answer "NO" and give them an opportunity to explain why they answered in this manner. Responses to this question tend to fall into two groups:

1) Refusals to pay because the air quality improvement has no value → a true $0 bid.
2) Responses along the lines that the respondent should not have to pay for the improvement in air quality, that someone else should be held responsible to pay. This can imply a non-true $0 bid value, or what we also consider as the rejection of a scenario built into a survey version.

Experimental evidence suggests that hypothetical markets, such as those set up by a CVM procedure, result in an excess number of zero bids. Therefore, there is some evidence for the need to further identify the nature of these excess “zeros” by some mechanism. One method is to identify individuals who give “NO” responses to the WTP/WTA questions, yet provide answers to previous questions, indicating that they do positively value the commodity. Such zero bids are the result of respondents rejecting the question’s scenario and do not necessarily represent true zero values. In the past, these responses have been considered distinct from “real” $0 bids that actually reflect the value an individual assigns to an improvement in the commodity of interest. These “NO” responses are typically identified by respondents’ comments such as:

- “I am unable to pay, or those who are more able to pay should do so,”
- “I feel that other interests are responsible and should pay the tab for improvements,”
- “I feel that other programs should be instituted first or other actions taken, before more money is to be spent.”

Another mechanism that can be utilized to trim excess zero bids from the sample involves the use of WTA values. A “NO” response in the WTA format is interpreted as “no amount of money would compensate for a loss in air quality”. As chapter 2 suggests, such refusals ($0 WTA bids), when accompanied by $0 WTP bids, can be interpreted as status quo bias. Thus, one can use a WTA refusal test to discard zero WTP values.

Base survey versions A and B were also designed to explore separate component values for health and visibility. Respondents were asked to break down the dollar values they provided in the WTP/WTA questions into percentage shares that they would allocate to health and visibility aspects (respectively) of clean air. This question format attempts to overcome the “superadditivity” problem which arises when people are asked to provide values for separate characteristics of a commodity and then give a total value for the commodity as a whole. In general, individuals are able to readily provide dollar values for the individual characteristics or components of a nonmarket commodity (whether it is health or visibility components of clean air, or use, existence, and option values for underground water). However, the summation of component values often exceeds an individual’s original total bid for the commodity. As discussed in the previous section, there is no single clear answer as to why this violation occurs. One interpretation may be that individuals actually envision the various aspects of a commodity as being inextricably linked as a joint product and that if they bid a value for one aspect they are necessarily obtaining other aspects that they see as being tied to the commodity. Whatever the reason, this format provides an alternative method to try to extract “reasonable” component values for different characteristics of the commodity.

Versions C and D The ability of respondents to separately value health and visibility is tested in survey variants C and D (figure 3.4), where these two versions mirror versions A and B, respectively, except for the format of the value question. Versions C and D ask respondents to value visibility and health individually and then provide the dollar amount which they would pay for the two characteristics combined. These variants should allow us to demonstrate that the superadditive approach yields inconsistent values and that an alternative methodology is required.

Version E This version (figure 3.5) employs some of the recommendations that Fischhoff and Furby (1988) and Mitchell and Carson (1989) have made in regard to the use of the CVM in surveys valuing air quality as part of their work for EPRI. The main difference between Version E and the other versions is that V-E presents its valuation question in a referendum format and contains additional context information. This was done first to help define the attributes of air quality, the causes of air pollution, reference and target levels for the quality of the air and certainty of provision of the change. Particular emphasis was made to distinctly state what past, current and future levels in air quality changes the respondents were required to consider before providing a value.

The second area of contrast was the formatting and wording of the WTP and WTA elicitation questions. Extra context was included to assist in the outlining of the payment procedure that this survey instrument would institute concerning the payments' attributes, the context (payment vehicle), who is to offer payment and the certainty of payment. The additional text and re-formatting of the valuation questions into a referendum format was designed to echo Mitchell and Carsons' work. The referendum format also attempts to address the concerns of Fischhoff and Furby associated with the social context of a transaction. Efforts were made to clarify what decision rule would be used to determine whether or not the air quality improvements would actually be made, and what other issues would be involved in the transaction.

A comparison of the results of V-E and V-A should show the impact of a substantial change in context on CVM survey values. We believe that, in the Denver metropolitan area, the value that people associate with an improvement in air quality may be somewhat "uncrystallized". The concept of crystallized attitudes in relation to question form, wording, and context in survey research was introduced by Schuman and Presser (1981). As discussed in Chapter 2, context effects can be significant in cases where values do not necessarily exist independently of the measuring instrument, or in other words, when values are uncrystallized to a significant degree.

Version F In Version F, respondents are asked to value shifts in a frequency distribution of air quality in terms of the number of days in a visual air quality category (e.g., "good", "typical" or “poor”) This version is designed to explore which scenario the public would value more: increasing the number of “good” air quality days or eliminating “poor” air quality days. All of the other variants (except for version H) used text to describe a one step shift in air quality for the average winter day with respect to the respondents own personal rating of such a day. However, this version presented changes in air quality using frequency distribution graphs.

Frequency distributions were constructed from fourteen observations of VAQ ratings that were obtained from the Stewart et. al. study (1983, 1984) for the Winter of 1981. Using OAQPS’s database for the Winters of 1981 and 1987 we constructed a “best predictor” model for visual air quality (VAQ). The estimated model was:

$$VAQ81 = 6.3858 - 0.2415*CO81 - 43.6048*S081 + 9.024*OZ81.$$

(t-values) (3.701) (-2.019) (-2.690) (0.179)

(F-value = 14.597)

VAQ was regressed on variables we thought to be highly correlated with visibility and have the highest predictive power for VAQ. We used CO, SO, and Ozone measurements for the days (and times) for which we had VAQ information. (Particulate measurements were excluded due to lack of readings from which to provide enough data points for a reasonable regression.) From this predictor model we constructed a frequency distribution for VAQ in 1981 (see Figure 3.6) and 1987 (see Figure 3.7). Note that these distributions show a substantial decrease in poor days and an increase in good days from 1981 to 1987 as measured by VAQ.

As the base for shifting frequency distributions in the survey, we simplified Figure 3.6 into Figure 3.8 by having only 3 categories of air quality and rounding frequencies. Improvements in the number of poor air quality days (Figure 3.9) and in good air quality days (Figure 3.10) were then constructed to be consistent with a one-half step increase in mean winter air quality.

Thus, the base case frequency distribution actually used in the survey showed the average number of days during Denver’s air pollution season that air quality fell in each category. Two scenarios followed which presented the different air quality frequency distributions of Figures 3.9 and 3.10. In the first WTP question, respondents compared the current distribution to an improved distribution in which the “very poor” days (a rating of 1 or 2) were eliminated, but the number of “very good” days (a rating of 6 or 7) remained the same. For the second WTP question, respondents gave WTP values for a different type of improvement in which the number of “very poor” days remained the same, but the number of “very good” days increased. By asking these two distinct WTP questions we were able to discern any differences which might exist in values for different types of air quality improvement, couched in term of changes in the frequency of certain types of days occurring. If such differences exist, this information may be useful for policy-making.

Additivity problems could be expected from this design. As discussed in the previous chapter, people tend to overvalue commodities when they think about them in a disaggregated way such as the number of days of each air quality type as opposed to an aggregate concept such as the average change in air quality. To overcome this, Version F tries to simplify the presentation as much as possible. However, given the manner in which means behave in response to shifts in frequency distributions, many people may not be able to link these means in a coherent fashion to the respective changes in the breakdown of days in each category of air quality. In addition, previous research has shown that individuals have difficulty interpreting frequency information.

Versions G and H The final two versions, V-G and V-H, were developed to provide a dramatic change to the "traditional" contingent valuation method of finding values for nonmarket commodities, and to explore the possibility of preference reversals. While Version G employs a traditional WTP value question format (see Figure 3.12), it differs from other versions in that respondents are asked to value improvements in two market commodities (a camera and stereo) as well as valuing an improvement in air quality. Respondents were asked to value an upgrade in quality from a less valuable to a more valuable camera or television. Just as in the WTP questions in the other survey versions, they were then asked to value a trade from poorer air quality to better air quality. This construct has the intention of providing a respondent with some practice in valuing goods that they are familiar with before considering the value for an improvement in air quality. A rationale for doing this is that while people are familiar with valuing private goods such as cameras or televisions, they are novices, for the most part, at valuing public goods such as air quality.

In contrast, version H asked respondents to choose either an improvement in air quality or an improvement in each of the private goods (see Figure 3.13). A comparison between V-G and V-H allows for a consistency check between valuation and choice response modes.

In both versions, respondents were shown photos of the commodities in question. These included two black and white photos of cameras, two black and white photos of televisions, and two color photos of the Denver area on a winter day. In version G, respondents answered WTP questions in the same manner as in all the surveys except they answered three value questions: for an upgrade in cameras, an upgrade in TV's, and an upgrade in air quality. In version H, respondents were given the three upgrade possibilities (upgrading a camera, a television set, or air quality) and asked to compare the improvement in air quality to the camera (Q-8), and television (Q-9) improvements, respectively. indicating which of the improvements their household would be willing to pay more for.

Pretest Description For Commodities Used In Versions G and H:

Twenty-five students from intermediate level economics classes were recruited to fill out a facsimile of version G with several different pairings of cameras, TV's and air quality photos. The visibility photos selected were A and C from the photo sheet and represented a one step difference on the seven step VAQ ladder. The mean dollar value for improvement in air quality also fell between the mean dollar value of improvement for the pairs of cameras and TV's used in the field survey. The mean dollar values and the standard deviations that were derived from the pre-test for each pairing used in the actual survey are:

Commodity	Mean $ Value for Improvements	Standard Deviation
Cameras	$42.40	39.61
TV's	77.60	63.16
Air Quality	66.60	68.98

Figure 3.2

Version A (base survey)

I. THE ISSUES

We would like to find out how Denver's air pollution and the Brown Cloud affect you and your household.

Q-1 How bothered have you been by the Brown Cloud's effect on what you can see in the distance (mountains, buildings, etc.)?

Not Bothered						Extremely Bothered
1	2	3	4	5	6	7

Q-2 Have you or your family been bothered by any health problems which you believe to be caused or aggravated by Denver's air pollution?

Not Bothered						Extremely Bothered
1	2	3	4	5	6	7

Q-3 How often have you read or heard about air pollution or the Brown Cloud? Please circle the appropriate response for each category.

1. Television	Never	Once or Twice	Many Times
2. Newspaper	Never	Once or Twice	Many Times
3. Radio	Never	Once or Twice	Many Times
4. Family	Never	Once or Twice	Many Times
5. Friends	Never	Once or Twice	Many Times
6. Other	Never	Once or Twice	Many Times

Please specify____

II. HOW DO YOU RATE DENVER'S AIR?

Q-4 Please refer to the visual air quality ladder on the photo sheet enclosed with this survey. The best possible visibility is rated as 7 and the lowest is rated as 1. Visibility, which is defined as the ability to see distant objects, improves as you move up the ladder. Using this ladder, and focusing on visibility alone (ignoring health effects for the moment) how do you rate the visual air quality in each of the photos A through F?

	Extremely Poor					Extremely Good	
1. Photo A	1	2	3	4	5	6	7
2. Photo B	1	2	3	4	5	6	7
3. Photo C	1	2	3	4	5	6	7
4. Photo D	1	2	3	4	5	6	7
5. Photo E	1	2	3	4	5	6	7
6. Photo F	1	2	3	4	5	6	7

Q-5 Again, using the visual air quality ladder to scale your answer, how would you rate the healthiness of the air in each of the photographs?

	Extremely Poor					Extremely Good	
1. Photo A	1	2	3	4	5	6	7
2. Photo B	1	2	3	4	5	6	7
3. Photo C	1	2	3	4	5	6	7
4. Photo D	1	2	3	4	5	6	7
5. Photo E	1	2	3	4	5	6	7
6. Photo F	1	2	3	4	5	6	7

Q-6 How would you rate the current overall air quality (thinking of both health and visibility) on an average winter day in the Denver metro area, using the visual air quality ladder to scale your answer?

	Extremely Poor					Extremely Good	
Denver's Average Winter Day:	1	2	3	4	5	6	7

III. THE VALUE OF CLEAN AIR TO YOU

Future extensions to the Better Air Campaign could reduce pollution from automobiles and tighten standards on industry. Catalytic converters or other forms of reducing smoke from wood burning stoves and fireplaces could become mandatory. Such actions could lead to higher prices for certain commodities (gasoline and/or wood-burning installations, for example).

Q-7 Health and visibility impacts often occur together when a brown cloud develops over the Denver metro area. *If we are going to have cleaner air we are all going to have to pay for it*. Assuming that everyone would have to pay, would you be willing to pay higher prices so that the average winter day's visual air quality increases by one step on the ladder (in comparison to your rating in Q-6)?

1. NO → WHY?

2. YES ↓

What is the MOST your household would be willing to pay EACH YEAR, in higher prices, to fund a one step increase in the average winter day's air quality on the air quality ladder for BOTH better visibility and fewer health problems from the Brown Cloud?

$___________

Q-8 Using the value you provided above (in Q-7), what percentage of that total would you attribute to better visibility and what percentage to fewer health problems?

a) percent due to visibility ______________________

b) percent due to health ______________________

On the other hand, air pollution could be allowed to get worse. There could be fewer restrictions on the fuels used for commuting and industry. The door could be left open for just about any type of industry to locate in the Denver metro area, and operations would be made less costly if emission controls were eased. Emission standards would also be relaxed on automobiles, resulting in less costly maintenance and repair bills. Such actions would lead to higher levels of pollution, however there could also be an offsetting decrease in prices as a result. We would like to know how much compensation is needed for you to accept more air pollution.

Q-9 If we spend less money on air pollution control the air quality will probably get worse. *In turn, we would all have more money, due to lower prices.* Would you be willing to accept lower prices as compensation for a decrease in overall air quality so that the average winter day's visual air quality decreases by one step on the ladder (in comparison to your rating in Q-6)?

1. NO → WHY?

2. YES ↓

What is the SMALLEST amount of money your household would be willing to accept EACH YEAR, in lower prices as compensation, for a one step decrease in the average winter day's air quality on the air quality ladder for BOTH worse visibility and more health problems from the Brown Cloud?

$____________

Q-10 Using the value you provided above (in Q-9), what percentage of that total would you attribute as compensation for a single step down on the visual air quality ladder due to worse visibility and what percentage to more health problems?

a) percent due to visibility ______________________

b) percent due to health ______________________

IV. ABOUT YOU

Q-11 Your age: ______ YEARS

Q-12 Your sex?

1. Female
2. Male

Q-13 Including yourself, how many members in your household are in each age group? (If none, write "0")

_____ Under 18 years of age
_____ 18-64
_____ 65 and over

Q-14 How much formal education have you completed?

1. No formal education
2. Some grade school
3. Completed grade school
4. Some high school
5. Completed high school
6. Trade school
7. Some college
8. Completed college
9. Some graduate work
10. Advanced college degree

Q-15 Are you presently: (circle number of the best answer)

1. Employed
2. Unemployed
3. Retired
4. Full-time homemaker
5. Student
6. Other__________

Q-16 What is your occupation?

JOB________________________

Q-17 What was the approximate annual gross income (before taxes) received last year by you and adult (18 years or older) family members living with you?

1. Under $5,000	6. $25,000-29,999	11. $60,000-69,999
2. $5,000-9,999	7. $30,000-34,999	12. $70,000-79,999
3. $10,000-14,999	8. $35,000-39,999	13. $80,000-89,999
4. $15,000-19,999	9. $40,000-49,999	14. $90,000-100,000
5. $20,000-24,999	10. $50,000-59,999	15. More than $100,000

Q-18 About how many total hours per week do you and other adult members of your household spend working at income generating employment?

_________ Hours

Q-19 What type of residence do you live in?

1. House
2. Apartment
3. Condominium/Townhouse
4. Studio
5. Other ______________________

Q-20 Do you own or rent your residence?

1. Own
2. Rent/lease

Is there anything we may have overlooked? Please use the space below for any additional comments you would like to make concerning Denver's air quality and its influence in your life.

Figure 3.3

Questions Unique to Versions B

II. HOW DO YOU RATE DENVER'S AIR?

The U.S. Environmental Protection Agency has set National Ambient Air Quality Standards at pollution levels designed to protect human health with an adequate margin of safety. Effects associated with elevated pollution levels include:

Carbon Monoxide:	Heart and circulatory problems such as aggravation of angina, and nervous system effects such as decreased alertness.
Sulfur Dioxide and Particulate Matter:	Aggravation of existing respiratory diseases such as asthma and bronchitis and increased respiratory symptoms such as shortness of breath and wheezing.
Ozone and Nitrogen Dioxide:	Respiratory symptoms such as coughing and chest pain, and risk of damage to the lungs.

The standards for several pollutants found in Denver's Brown Cloud are sometimes exceeded on bad days in the winter months. Some of these pollutants may impair visibility, while others (carbon monoxide and ozone) are colorless gasses. The "Air Pollution Index" announced each day shows how the air quality compares to the Standard. For instance, an Index of 180 for Carbon Monoxide means that the carbon monoxide levels are 80% higher than the Standard. The photos on the enclosed sheet were taken on days with the following Air Pollution Indices:

PHOTO	CARBON MONOXIDE	SULFUR DIOXIDE	OZONE
Example Step 2	101	186	2
Example Step 6	25	0	23
A	118	136	1
B	44	46	7
C	62	110	7
D	36	13	17
E	115	67	1
F	115	67	1

Figure 3.4

Questions Unique to Versions C and D

The following question pertains ***only*** to ***visibility*** impacts of air pollution and the Brown Cloud; health impacts will be considered next.

Q-7 The color of the air, the ability to see the Denver skyline and its mountain backdrop can vary from time to time, particularly across seasons. *If we are going to have cleaner air we are all going to have to pay for it*. Assuming that everyone would have to pay, would you be willing to pay higher prices so that the average winter day's visual air quality increases by one step on the ladder (in comparison to your rating in Q-6)?

1. No → WHY?

2. YES ↓

What is the MOST your household would be willing to pay EACH YEAR, in higher prices, to fund a one step increase in the average winter day's air quality on the air quality ladder for better visibility?

$___________

The following question pertains ***only*** to ***health*** aspects of the Brown Cloud.

Figure 3.4 (con't)

Q-8 Air pollution may affect some peoples' health. *Once again, if we are going to have cleaner air we are all going to have to pay for it.* Assuming that everyone would have to pay, would you be willing to pay higher prices so that the average winter day's air quality increases by one step on the ladder (in comparison to your rating in Q-6) just to improve health?

1. No → WHY?

2. YES ↓

What is the MOST your household would be willing to pay EACH YEAR, in higher prices, to fund a one step increase in the average winter day's air quality on the air quality ladder for fewer health problems from the Brown Cloud?

$___________

The following question asks about your total value for improving both health and visibility in the Denver metro area.

Q-9 Health and visibility impacts often occur together when a brown cloud develops over the Denver metro area. Again, assuming that everyone would have to pay, would you be willing to pay higher prices so that the average winter day's visual air quality increases by one step on the ladder (in comparison to your rating in Q-6)?

1. No → WHY?

2. YES ↓

What is the MOST your household would be willing to pay EACH YEAR, in higher prices, to fund a one step increase in the average winter day's air quality on the air quality ladder for BOTH better visibility and fewer health problems from the Brown Cloud?

$___________

Figure 3.5

Questions Unique to Version E

Several programs have been tried in the Denver area, including the voluntary Better Air Campaign and the mandatory oxygenated fuels program. Last year, the Better Air Campaign reduced wood burning by 7 percent. The oxy-fuels program which uses special gasoline also cut carbon monoxide emissions by 9.5 percent. These programs have resulted in Denver being dropped from first to seventh place on the Environmental Protection Agency's list of worst cities for carbon monoxide pollution for Winter 1987-88.

Q-8 In your opinion, if Denver does nothing further about air pollution and the Brown Cloud, over the next 5-10 years, the problem is likely to

1. get much worse.
2. get somewhat worse.
3. stay about the same.
4. get somewhat better.
5. get much better.

The Better Air Campaign currently runs for three months, November through January; it encourages commuters not to drive one day a week and on high pollution days, and it also promotes wood-burning bans on high pollution days. The mandated high-oxygen fuels program, however has increased the oxygen content of blended fuels from 1.5 percent to 2 percent and is being expanded by one month, to operate from November through February.

Q-9 Among the following actions, which have you and your household taken or intend to take? (Please circle all that apply.)

1. Reduced the use of your fireplace and/or wood-burning stove.
2. Taken some other form of transportation (i.e. bus, bicycle, walk, etc.) rather than your car to work or shopping.
3. Car-pooled.
4. Used oxygenated gasoline.
5. Did not drive one day per week (to work or other places).
6. Other (please specify)________________________

Figure 3.5 (Con't)

Q-10 To what degree do you think the following causes contribute to the air pollution problem in the Denver metro area?

	Very Low Contribution					Very High Contribution	
1. Automobiles	1	2	3	4	5	6	7
2. Fireplaces	1	2	3	4	5	6	7
3. Power plants	1	2	3	4	5	6	7
4. Industry	1	2	3	4	5	6	7
5. Weather Conditions	1	2	3	4	5	6	7
6. Other (please specify)	1	2	3	4	5	6	7

____

Q-11 How long have you lived in the Denver metro area?

____________ YEARS ____________ MONTHS

Q-12 How do you feel about the amount of attention given to the air pollution problems of the Denver metro area?

TOO LITTLE					TOO MUCH	
1	2	3	4	5	6	7

III. THE VALUE OF CLEAN AIR TO YOU

Future extensions of the Better Air Campaign could also include tighter emission standards on automobiles and tighter standards on industrial pollution. Catalytic converters or other forms of filtering smoke from wood burning stoves and fireplaces could become mandatory. Such actions could lead to higher prices for certain commodities (gasoline and/or wood-burning installations, for example) or higher taxes (for the enforcement of tighter standards and stronger regulations). Moreover, companies pay for the cost of air pollution controls imposed upon them, which increases the cost of doing business, in turn passing on some or most of these costs into the prices paid for their products.

Figure 3.5 (Con't)

Q-13 In your opinion, if Denver implements a tough program for air pollution and the Brown Cloud, over the next 5-10 years, the problem is likely to

1. get much worse.
2. get somewhat worse.
3. stay about the same.
4. get somewhat better.
5. get much better.

Q-14 How would you rate the current overall air quality (thinking of both health and visibility) on an average winter day in the Denver metro area, using the visual air quality ladder to scale your answer?

	Extremely Poor				Extremely Good		
Denver's Average Winter Day:	1	2	3	4	5	6	7

Imagine that a referendum for a tax increase on gasoline and diesel fuel is proposed in the Denver metro area. If the referendum is passed, everyone would be paying higher prices as a result of higher taxes in order to fund the programs it would support. The money would be used ***only*** for the improvement of health and visibility, and no other purpose. *At the moment we don't know what everyone would have to pay to cover all the costs of air pollution control programs, so we need to find out how much it is worth to people.*

Q-15 The impacts on health and visibility often occur together when a Brown Cloud develops over the Denver metro area. Would you consider voting for a referendum, *if the improvement could be guaranteed*, which would require you to pay higher taxes and/or prices to improve overall air quality, so that the average winter day's visual air quality increases by one step on the ladder (as compared to your rating in Q-14)?

Figure 3.5 (Con't)

1. NO → WHY?

2. YES ↓

What is the MOST your household is willing to pay EACH YEAR, in higher taxes and/or prices, before you would vote NO on such a referendum, so there would be a one step increase in the average winter day's air quality on the visual air quality ladder for BOTH visibility and health effects of air pollution and the Brown Cloud?

$____________

The amount you indicate will tell us what it is ***really worth*** *to your household to get such an improvement. If the improvement actually costs less than people are willing to pay, you would only have to pay what it would cost. If the gain turned out to cost more than people are willing to pay, it would not be carried out.*

On the other hand, air pollution standards could be lowered. There could be fewer restrictions on the fuels used for commuting and industry. The door could be left open for most any type of industry to locate in the Denver metro area, and operations could be made less costly if emission controls were eased. Emission standards could also be relaxed on automobiles, resulting in less costly maintenance and repair bills. Such actions would lead to higher levels of pollution.

Imagine that a referendum for a tax decrease on gasoline and diesel fuel is proposed in the Denver metro area. If the referendum is passed, everyone would be paying lower prices as a result, and lower taxes in order to fund the programs it would support. Only the programs for the improvement of health and visibility of Denver's air would be cut back, and no others.

Figure 3.5 (Con't)

Q-16 The impacts on health and visibility often occur together when a Brown Cloud develops over the Denver metro area. Would you consider voting for a referendum which would require you to pay lower taxes and/or prices but decrease overall air quality so that the average winter day's visual air quality decreases by one step on the ladder?

1. NO → WHY?

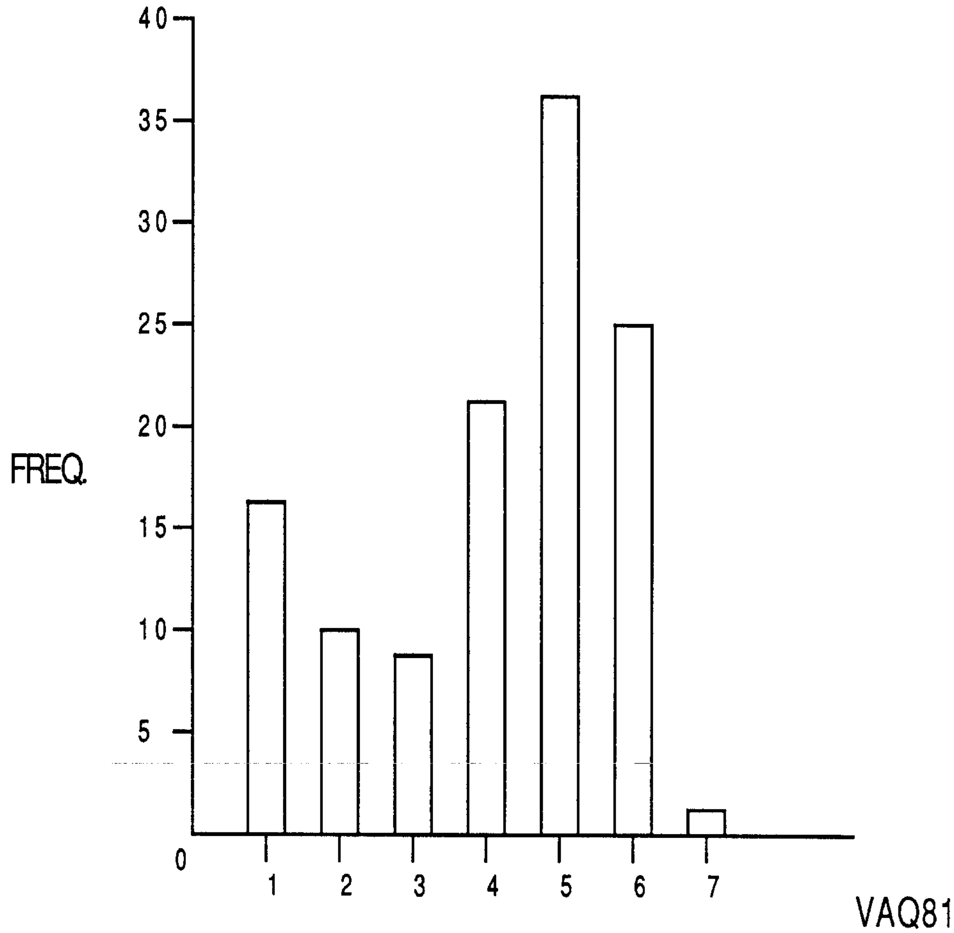
2. YES ↓

What is the LEAST your household is willing to accept EACH YEAR, in lower taxes and/or prices, before you would vote YES on such a referendum, so there would be a one step decrease in the average winter day's air quality on the visual air quality ladder for BOTH visibility and health effects of air pollution and the Brown Cloud?

$____________

Figure 3.6

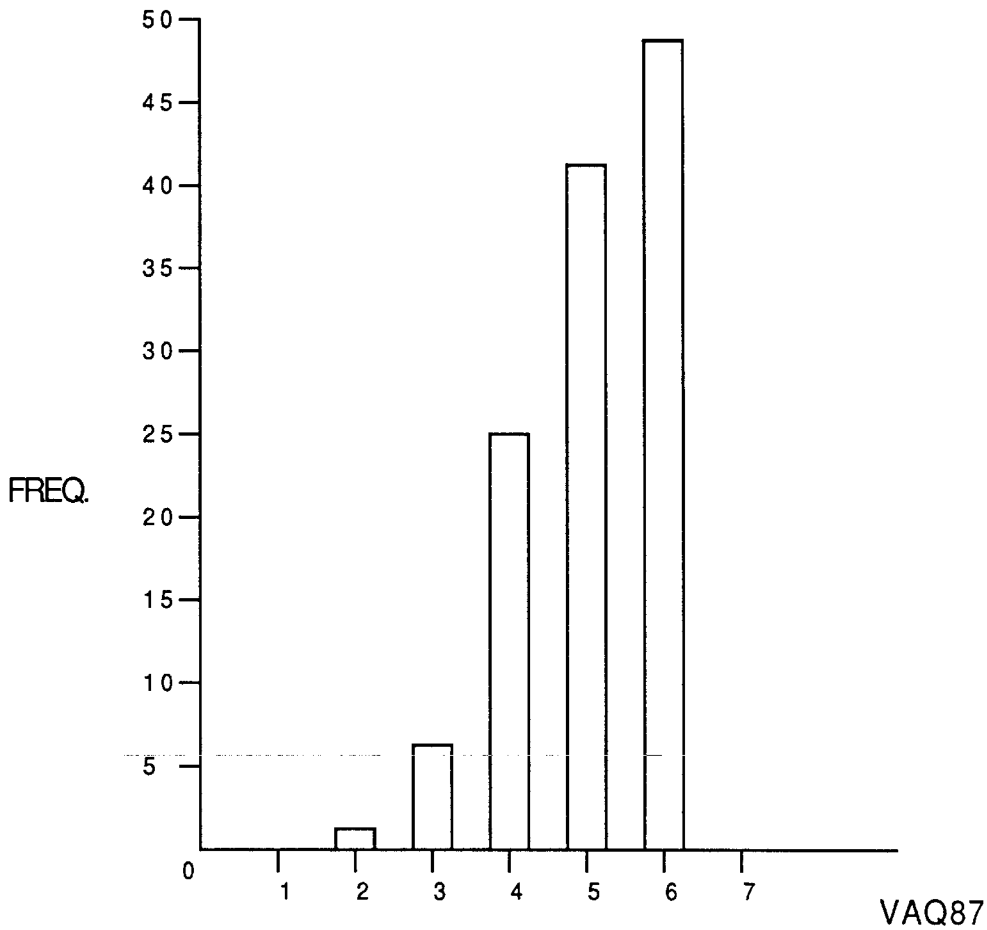
Frequency Bar Chart
VAQ 1981 (Midpoint)



VAQ	FREQUENCY	CUMULATIVE FREQUENCY	PERCENT	CUMULATIVE PERCENT
1	16	16	13.22	13.22
2	10	26	8.26	21.49
3	8	34	6.61	28.10
4	23	57	19.01	47.11
5	37	94	30.58	77.69
6	25	119	20.66	98.35
7	2	121	1.65	100.00

Figure 3.7

Frequency Bar Chart
VAQ 1987 (Midpoint)



VAQ	FREQUENCY	CUMULATIVE FREQUENCY	PERCENT	CUMULATIVE PERCENT
1	0	0	0.00	0.00
2	1	1	0.83	0.83
3	6	7	4.96	5.79
4	25	32	20.66	26.45
5	41	73	33.88	60.33
6	48	121	39.67	100.00
7	0	121	0.00	100.00

Figure 3.8

Version F Air Quality Frequency Distributions

(Base Case)

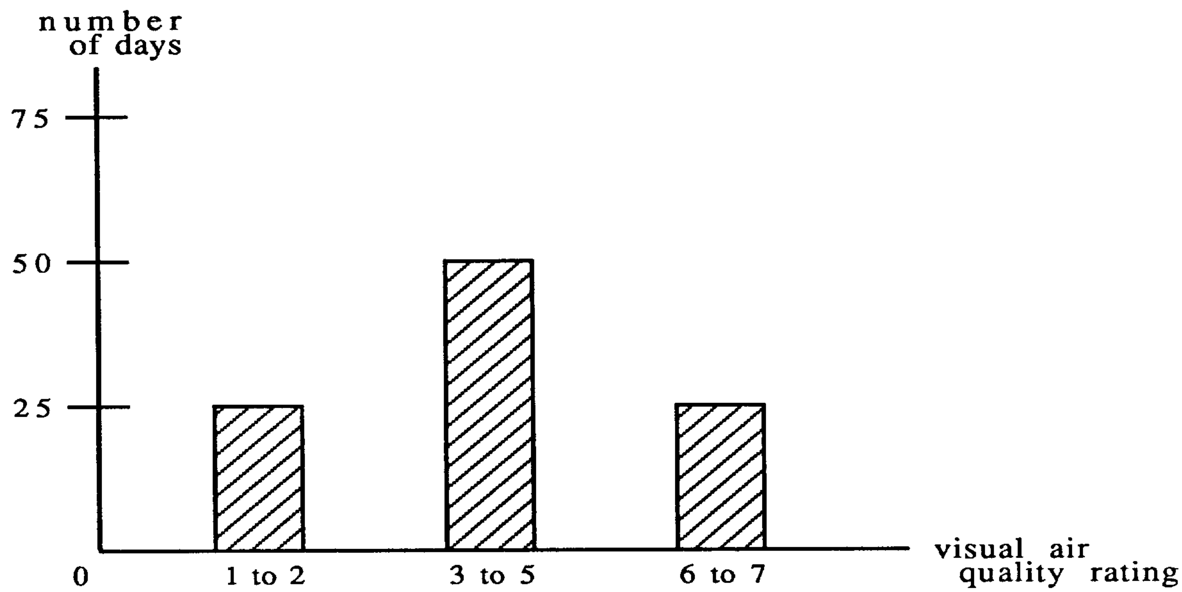


Figure 3.9

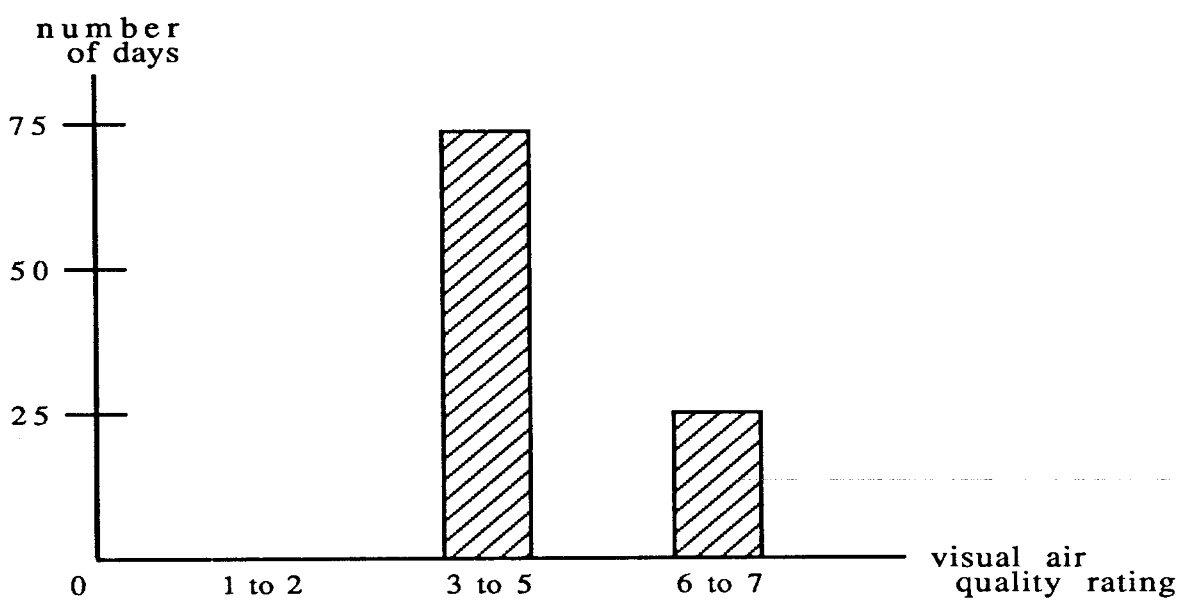


Figure 3.10

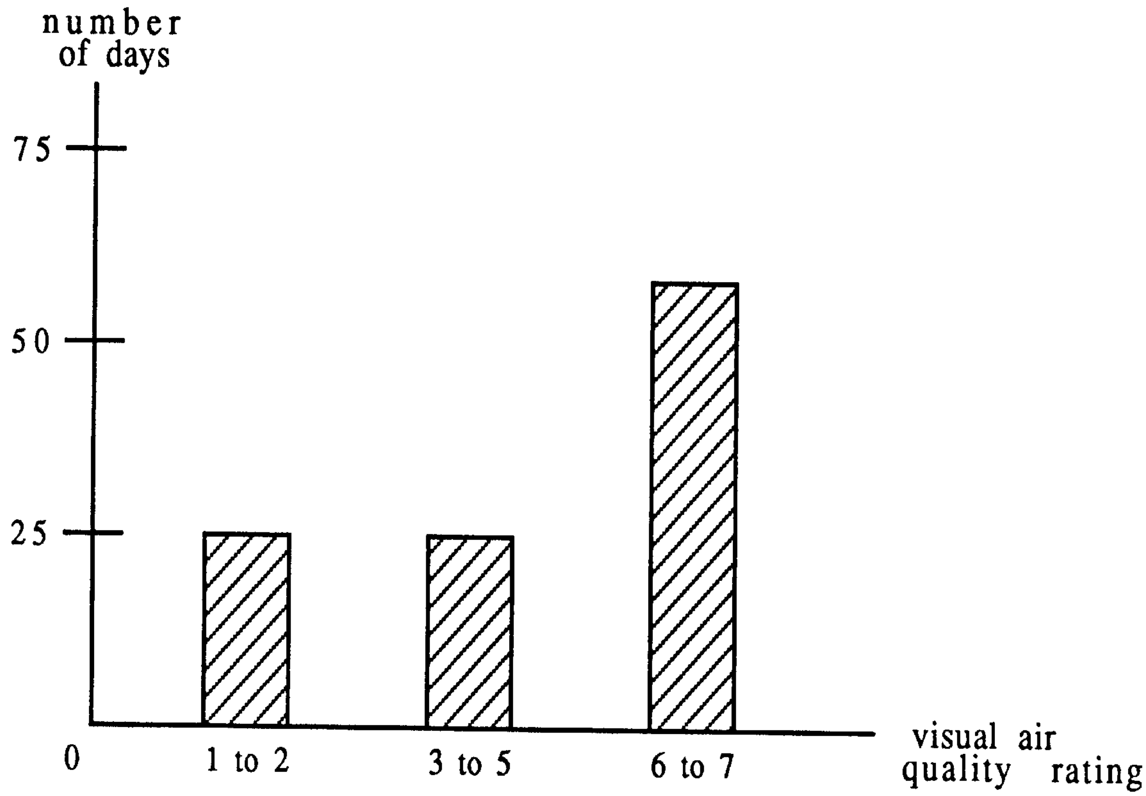

number
of days
75
50
25
0
1 to 2
3 to 5
6 to 7
visual air
quality rating

Figure 3.11

III. THE VALUE OF CLEAN AIR TO YOU

In this section we will ask you about the value of improving Denver's air. However, before doing this we would like you to think about the value of improving two household goods.

Q-7 On the previous page you will see photographs of two cameras. The camera on the top (1A) is a 35 mm camera, it has automatic flash and automatic focus (instead of fixed focus). The other camera is even more advanced. It allows you to use whatever kinds of lenses you would like, has an advanced autofocus system, and a programmed electronic shutter.

Now suppose you own a new camera like the one shown in Picture 1A. Would your household be willing to trade it for a new camera as shown in Picture 1B, if you also had to pay some extra money in the trade? Please answer below.

1. NO → WHY?

2. YES ↓

What is the MOST your household would be willing to pay to trade in the camera in picture 1A for the camera in picture 1B?

$____________

Figure 3.11 (con't)

Q-8 The pictures on the previous page depict two television sets. The T.V. on the top (2A) is a color TV with no particular extras. The T.V. on the bottom (2B) has high contrast images, a sleep timer, and a remote control that activates all T.V. functions.

Now suppose you own a new T.V. like the one shown in Picture 2A. Would your household be willing to trade it for a new T.V. as shown in Picture 2B, if you also had to pay some extra money in the trade? Please answer below.

1. NO → WHY?

2. YES ↓

What is the MOST your household would be willing to pay to trade in the television in picture 2A for the television in picture 2B?

$____________

Figure 3.11 (con't)

Q-9 Now please compare photographs 3A and 3B. They are both pictures of a winter day. One of them shows a greater amount of visible pollution than the other. Suppose 3A depicts the amount of pollution that you live with most days of the winter.

Would you be willing to pay higher prices (for example, gasoline prices) to improve overall air quality--so that your average winter day improves from being like Picture 3A to being like Picture 3B? Please answer below.

1. NO → WHY?

2. YES ↓

What is the MOST your household would be willing to pay EACH YEAR, in higher prices, to fund the improvement in average winter-day air quality from that shown in picture 3A to that shown in picture 3B?

$____________

Figure 3.12

III. THE VALUE OF CLEAN AIR TO YOU

On the previous page you see two photographs of Denver's brown cloud. Picture 1B shows cleaner air than does Picture 1A. Consider both the health and visibility effects if the average winter day of air quality improved from that in Picture 1A to that in Picture 1B. Such an improvement would probably cost you money each year, in increased prices.

Now look to the next page, where you will find photographs of two cameras. The camera on the top (2A) is a 35 mm camera that is very automatic, but does not have any extras that could help it take exceptional pictures. The camera on the bottom (2B) is more advanced. It allows you to use whatever kinds of lenses you would like, has an advanced autofocus system, and a programmed electronic shutter.

Now suppose you own a new camera like the one shown in Picture 2A. Suppose you could trade it for a new camera as shown in Picture 2B. This improvement would probably cost you money.

In answering the next question, think about the value your household would place on a year's improvement in air quality (both health and visibility) as represented by going from Picture 1A to Picture 1B for the average winter day. DO NOT consider the value such an improvement would have to OTHER PEOPLE. Compare the MOST your household would be willing to pay for this improvement for ONE YEAR, in higher prices, with the MOST your household would be willing to pay to trade in camera 2A for camera 2B.

Q-8 Which improvement would your household be willing to pay MORE for? Circle one of the three answers below.

1. More for one year's improvement in air quality, from Picture 1A to Picture 1B.
2. The same.
3. More for the improvement in cameras, from Picture 2A to Picture 2B.

The next question asks you for the same kind of comparison. Again consider your household's willingness to pay for one year's improvement in air quality as represented by the changes from Picture 1A to Picture 1B. Remember to consider the value of the improvement only to your household, not to other people.

This time, for the comparison, look at the television sets pictured on the next page. The T.V. on the top (3A) is a color TV with no particular extras. The T.V. on the bottom (3B) has high contrast images, a sleep timer, and a remote control that performs all T.V. functions.

Figure 3.12 (Con't)

Now suppose you own a new T.V. like the one shown in Picture 3A. Suppose you could trade it for a new T.V. as shown in Picture 3B, if you also had to pay some extra money in the trade.

Q-9 Which improvement would your household be willing to pay MORE for? Circle one of the three answers below.

1. More for the one year's improvement in air quality.
2. The same.
3. More for the improvement in T.V.'s.

CHAPTER 4
Descriptive Statistics

4.1 Introduction

This chapter will discuss the descriptive results from the field experiment that implemented contingent valuation methods. First, a description of the mailing procedures, the population sampled, and the response rate shall be covered. Results from sections I, II and IV of the surveys which are the same across variants, are pooled and reported. Results for questions that were idiosyncratic to version E will be presented separately. All descriptive findings pertaining to CVM value questions (section III of the surveys) are covered in Chapter 5 along with the data analysis.

4.2 Sample Design

The mailing procedures and constitution of the sample population used in the implementation of this study were designed following the Dillman Total Design Method (TDM). The TDM procedure aims to maximize response rates through specific design and implementation strategies. For example, personalizing the mailing to include a cover page, hand-signed cover letter, hand-stamped envelope, and follow-up postcard are key factors for encouraging responses. Examples of these are presented in figures 4.1-4.5. The cover letter (Figure 4.1) is designed to introduce the respondent to the topic of the questionnaire. It describes in general what the survey is about, who should fill it out and who is conducting the research. Inside the cover of the survey is a second statement of what the survey is about, and who should fill it out. Figures 4.2 and 4.3 present the reminder postcard and follow-up letter, respectively.

According to the TDM, the survey must be printed and folded into a booklet measuring 8 inches by 6 inches. The surveys that we sent were from 6 to 9 pages long, depending on the version sent, including the cover and the space for comments. One version of the survey, a cover letter, and a self-addressed, stamped envelope were mailed to each individual in the survey.

The mailout package also included a two dollar bill. This was enclosed to encourage the household to fill out and return the survey. This monetary incentive is not part of TDM, but we have found it increases response rates significantly.

The households surveyed constitute a random sample from the Denver Metro area. Two hundred households were selected from the city of Denver and four of its neighboring suburbs of Arvada, Aurora, Littleton, and Wheat Ridge, for *each* of the eight survey variants for a total sample size of 1600.

In an effort to maximize response rates, the TDM specifies follow-up mailing procedures. One week after initial mailing, a postcard was sent reminding respondents of the importance of completing and returning the survey. If a response was not received after three weeks, a second survey, cover letter and self-addressed, stamped envelope were sent out. However, this second mailing package did not contain another two dollar bill. This approach yielded an overall response rate from the first mailing and postcard of 71% — see table 4.1 (response rate table).

Table 4.1

Response Rates

Day	Responses	Bad Addresses
1	188	50
2	159	93
3	90	18
4	55	17
5	42	20
6	66	7
7	86	7
8	54	1
9	28	1
10	17	3
11	14	0
12	8	1
13	11	1
14	7	1
15+	23	0
TOTAL:	848	220
% of TOTAL: (out of 1600)	53%	15.75%

Version	A	B	C	D	E	F	G	H
Number of responses:	122	108	118	109	112	121	122	126
Number of bad addresses:	32	41	35	44	36	30	36	25
Response Rates (%):	72.6	67.9	71.5	69.9	68.3	71.2	74.4	72.0

Total Number of Responses = 938

Total Number of Bad Addresses = 279

Overall Response Rate = 71%

Figure 4.1

Date, 1989

The Brown Cloud air pollution problem in the Denver Metro Area is an issue of increasing concern. However, little is known about what people think about air pollution and how they respond to the impacts it may have in their lives. In order to better assess what should be done about air pollution in the Denver area, we need the benefit of your experience.

You are one of a small number of households who are being asked what they think about Denver's Brown Cloud. Your name was drawn randomly from a list of households in the Denver Metro Area. In order for the results to truly represent the opinions and experience of those people who live in the Denver area, it is important that each questionnaire be completed and returned. It will take you about 15 minutes. Your answers will be combined with others in the city of Denver and its surrounding suburbs to form a profile of the area's views of air pollution issues.

Since this questionnaire asks specifically about your household's perceptions of Denver's air pollution, we ask that it be filled out by an adult in your household. You can be assured of complete confidentiality. In fact, your name will never be associated with this information. The number on the questionnaire is only so your name can be checked off the list when it is returned.

Since your responses are so valuable to us, we enclose $2 for your time and effort.

If you would like to receive a free summary of the survey results, please write "send results" on the back of the return envelope.

Many thanks for your help with this important effort.

Sincerely,

Gary McClelland
Project Director

Figure 4.2

April 25, 1989

Last week a questionnaire was mailed to you seeking information which is crucial in evaluating what people think about the air pollution issues of the Denver area.

If you have already completed and returned the questionnaire, accept our sincere thanks. If not, please do so today. Denver's Brown Cloud is a matter of increasing concern. Therefore, it is extremely important that your answers also be included in the study.

If by some chance you did not receive the questionnaire, or it was misplaced, please write us and we will get another one in the mail to you immediately.

Sincerely,

Gary McClelland
Project Director

Figure 4.3

Date, 1989

I am writing to you about our study of what people think about Denver's air pollution problem. To date, we have not yet received your completed questionnaire.

The large number of questionnaires returned is very encouraging. But whether we will be able to describe accurately what people think about the Brown Cloud in the Denver area depends on you and the others who have not yet responded. This is because our past experiences suggest that those of you who have not yet sent in your questionnaire may have very different opinions compared to those who have responded.

This study has been undertaken in the belief that people's attitudes towards air pollution should be incorporated into public management policies. Your opinions will be extremely valuable towards evaluating the worth of such programs. The usefulness of our results depends on how accurately we are able to describe the perceptions of the people in the Denver area.

In case our previous correspondence did not reach an adult in your household whose response is needed, a replacement questionnaire is enclosed. We urge you to complete and return it as quickly as possible.

We'll be happy to send you a copy of the results if you want one. Simply put your name, address, and "copy of results requested" on the back of the return envelope.

Your contribution to the success of this study will be appreciated greatly.

Most sincerely,

Gary McClelland
Project Director

4.3 Descriptive Statistics Pooled Across All Survey Variants

This discussion will follow the order of the facsimile presentation in Figure 4.1 which can be found at the end of this section. The first two questions in all variations asked how bothered the respondent had been by Denver's air pollution and its Brown Cloud. This was done first in relation to visibility (Q-1) and then in relation to health impacts (Q-2). The vast majority of respondents indicated that they have been bothered or extremely bothered by the Brown Cloud's effects on visibility. Over 74% of the sample rated it as a 5, 6, or 7 on the seven point scale where a seven indicates "extremely bothered". In contrast, it is interesting to observe that over 55% of the same sample indicated that they were "not bothered" by air pollution health problems; circling 1, 2, or 3 on the scale. This result leads to some interesting comparisons in the analysis of values for visibility and health effects in the next chapter, (i.e. health often appears to be valued more highly than visual air quality).

The final question of Section I (with the exception of V-E) provides an indication of how informed our sample of households was about air pollution in the Denver area. It is clear from their responses that a large majority of the respondents have heard, read and talked about air pollution or the Brown Cloud many times. This was not surprising since Denver's air pollution issues are a frequent matter of public debate throughout Colorado. For example, during this study, a Denver metro area paper carried the headline, "Governor Builds $1.7 Million Dollar War Chest For Brown Cloud" along with associated articles during the week of May 21, 1989.

Section III contains three rating exercises (once again with the exception of V-E, which has additional questions that will be presented later in this chapter). Questions (Q-4) and (Q-5) on the facsimile of Figure 4.1 provide the ratings respondents gave for the corresponding photos on the color insert sheet. In comparison, these photos have ratings, from the earlier VAQ study conducted by the National Center for Atmospheric Research (see Stewart et. al., 1983, 1984), of:

PHOTO	VAQ
A	3
B	7
C	2
D	6
E	3
F	4
Example step 6	6
Example step 2	2

Given the experience obtained in rating VAQ, respondents are then asked in (Q-6) to rate Denver's average winter day. Note that, in versions A, B, C, D and E respondents are asked to *value* a one step movement along the air quality ladder with respect to the rating they gave to Denver's average winter day on the seven step scale. Of the 923 responses to this question, approximately 42% provided a rating of 3 and 26% provided a rating of 4, which accounts for almost 68% of the total response. The mean ranking of the responses was 3.3 with a standard deviation of 1.02.

The final section, Section IV, consists of socioeconomic and demographic questions. The mean age of the respondents was 41.7 years while the most frequent age given by respondents fell into the 30 to 40 year old age bracket. The majority of persons completing the surveys were male (71.2%). An average household consisted of 1.3 "persons" under the age of 18, 1.9 persons between 18 and 64, and 0.32 over the age of 64. More than 84% of those surveyed had formal education beyond the high school level, with 52.9% having some college or completed college and 24.6% having gone on to some graduate work or earned an advanced college degree.

Seventy-six percent (76.4%) of the respondents were employed while another 18.2% were either retired, full-time homemakers or students, and 5.3% categorized themselves as unemployed or other. The mean total hours of labor per week per household was 59.6 hours/week, with 40 hours/week being the most frequent response.

Finally, the vast majority of respondents, 84.6%, lived in houses as opposed to other types of residences. Eighty-five percent owned their homes, while 15.5% said they rented or leased.

The annual gross income (before taxes) for over 50% of the households was between $30,000 and $59,000 per year with a median income of approximately $44,000. This figure is significantly higher than actual median household incomes obtained from census data in our sample area of $35,233. Although several factors may account for this discrepancy between our survey results and adjusted census data, the most important involves the survey implementation process.

In compliance with University of Colorado standards, the physical implementation of the survey was subcontracted by the University to the LOWEST bidding survey research firm. This firm was negligent in obtaining accurate, up to date mailing lists which produced a large number of bad addresses (13.75%). During analysis of these undeliverable surveys, it was discovered that the distribution of bad addresses was not exactly random. Most of the bad addresses were residents of apartment buildings and trailer parks (the incorrect address usually being attributable to missing apartment numbers, trailer unit number, etc.) Unfortunately, the undersampling of the "apartment community" was determined post facto.

It is conjectured that residents of apartment buildings and trailer parks typically have lower median incomes than homeowners. (Homeowners consisted of almost 85% of our sample.). Consequently, over-sampling homeowners produced the higher median household income values. To correct for this sampling bias, it is possible to perform a weighted regression where low income bids in the sample are overweighted in regression analyses, in effort to match general population demographics.

Figure 4.4

Facsimile of Descriptive
Statistics Pooled Across All Survey Versions

(Perc. Freq. = Percent of Frequency of Occurrence)

I. THE ISSUES

We would like to find out how Denver's air pollution and the Brown Cloud affect you and your household.

Q-1 How bothered have you been by the Brown Cloud's effect on what you can see in the distance (mountains, buildings, etc.)?

	Not Bothered						Extremely Bothered		
	1	2	3	4	5	6	7	Mean	5.26
(Perc. Freq.)	(1.2)	(3.9)	(6.3)	(14.5)	(26.5)	(26.0)	(21.7)	Std. Dev.	1.41
								n = 933	

Q-2 Have you or your family been bothered by any health problems which you believe to be caused or aggravated by Denver's air pollution?

	Not Bothered						Extremely Bothered		
	1	2	3	4	5	6	7	Mean	3.37
(Perc. Freq.)	(20.7)	(19.3)	(14.1)	(14.7)	(15.8)	(10.3)	(5.2)	Std. Dev.	1.85
								n = 932	

Q-3 How often have you read or heard about air pollution or the Brown Cloud?

Please circle the appropriate response for each category.

(Perc. Freq.)			
1. Television	Never (0.2)	Once or Twice (5.4)	Many Times (94.4)
2. Newspaper	Never (2.2)	Once or Twice (16.2)	Many Times (81.6)
3. Radio	Never (3.5)	Once or Twice (20.4)	Many Times (76.1)
4. Family	Never (6.1)	Once or Twice (32.0)	Many Times (61.9)
5. Friends	Never (7.5)	Once or Twice (28.0)	Many Times (64.5)
6. Other	Never (26.0)	Once or Twice (19.5)	Many Times (54.5)

Please specify____

II. HOW DO YOU RATE DENVER'S AIR?

Q-4 Please refer to the visual air quality ladder on the photo sheet enclosed with this survey. The best possible visibility is rated as 7 and the lowest is rated as 1. Visibility, which is defined as the ability to see distant objects, improves as you move up the ladder. Using this ladder, and focusing on visibility alone (ignoring health effects for the moment) how do you rate the visual air quality in each of the photos A through F?

(Perc. Freq.)	Extremely Poor 1	2	3	4	5	6	Extremely Good 7	Mean	Std. Dev.
1. Photo A n = 907	(5.1)	(20.7)	(40.5)	(25.8)	(6.6)	(1.0)	(0.3)	3.12	1.03
2. Photo B n = 900	(0.6)	(1.8)	(3.6)	(6.4)	(22.7)	(48.4)	(16.6)	5.60	1.11

Q-4 (Con't)

(Perc. Freq.)	Extremely Poor						Extremely Good	Mean	Std. Dev.
3. Photo C	1	2	3	4	5	6	7		
n = 907	(81.7)	(13.1)	(1.7)	(0.3)	(0.2)	(0.7)	(2.3)	1.35	1.06
4. Photo D	1	2	3	4	5	6	7		
n = 898	(1.4)	(0.9)	(0.3)	(0.9)	(2.0)	(19.6)	(74.8)	6.59	0.99
5. Photo E	1	2	3	4	5	6	7		
n = 900	(31.2)	(48.8)	(12.2)	(1.9)	(1.6)	(3.3)	(1.0)	2.08	1.19
6. Photo F	1	2	3	4	5	6	7		
n = 901	(4.0)	(20.2)	(37.2)	(24.1)	(12.1)	(2.1)	(0.3)	3.28	1.12

Q-5 Again, using the visual air quality ladder to scale your answer, how would you rate the healthiness of the air in each of the photographs?

(Perc. Freq)	Extremely Poor						Extremely Good	Mean	Std. Dev.
1. Photo A	1	2	3	4	5	6	7		
n = 899	(10.2)	(26.9)	(33.8)	(21.0)	(5.6)	(1.9)	(0.6)	2.93	1.17
2. Photo B	1	2	3	4	5	6	7		
n = 897	(1.0)	(1.9)	(4.5)	(10.9)	(24.5)	(42.1)	(15.1)	5.43	1.21
3. Photo C	1	2	3	4	5	6	7		
n = 899	(77.8)	(14.5)	(3.3)	(1.3)	(0.4)	(1.3)	(1.3)	1.42	1.05
4. Photo D	1	2	3	4	5	6	7		
n = 890	(1.3)	(1.1)	(1.2)	(2.2)	(8.1)	(27.0)	(59.0)	6.31	1.14
5. Photo E	1	2	3	4	5	6	7		
n = 898	(39.3)	(40.1)	(12.6)	(3.2)	(1.9)	(2.1)	(0.8)	1.98	1.16
6. Photo F	1	2	3	4	5	6	7		
n = 892	(8.5)	(23.8)	(33.7)	(20.9)	(7.8)	(4.3)	(1.0)	3.13	1.28

Q-6 How would you rate the current overall air quality (thinking of both health and visibility) on an average winter day in the Denver metro area, using the visual air quality ladder to scale your answer?

	Extremely Poor						Extremely Good
Denver's Average Winter Day:	1	2	3	4	5	6	7
(Perc. Freq)	(2.7)	(17.0)	(41.5)	(25.7)	(11.6)	(1.3)	(0.1)

n = 923

Mean: 3.3

Std. Dev.: 1.02

III. THE VALUE OF CLEAN AIR TO YOU

(Covered in Chapter 6)

IV. ABOUT YOU

Q-11 Your age: ______ YEARS

Age Bracket	Perc. Freq.		
under 20	1.1	n = 930	
20 - 30	12.8	Mean	41.68
30 - 40	40.4	Std. Dev.	13.16
40 - 50	19.7		
50 - 60	13.8		
60 - 70	9.2		
over 70	3.0		

Q-12 Your sex?

Perc. Freq.		
1. Female	28.4	n = 932
2. Male	71.2	

Q-13 Including yourself, how many members in your household are in each age group? (If none, write "0")

	Mean	Std. Dev.	n
_____ Under 18 years of age	1.28	1.15	688
_____ 18-64	1.89	0.82	893
_____ 65 and over	0.32	0.67	417

Q-14 How much formal education have you completed?

	Perc. Freq.		Perc. Freq.
1. No formal education	0.1	6. Trade school	7.1
2. Some grade school	0.1	7. Some college	27.3
3. Completed grade school	0.4	8. Completed college	25.6
4. Some high school	1.7	9. Some graduate work	10.4
5. Completed high school	12.6	10. Advanced college degree	14.2

n = 927

Q-15 Are you presently: (circle number of the best answer)

	Perc. Freq.		Perc. Freq.
1. Employed	76.4	4. Full-time homemaker	5.8
2. Unemployed	2.3	5. Student	1.7
3. Retired	10.7	6. Other__________	2.8

n = 929

Q-17 What was the approximate annual gross income (before taxes) received last year by you and adult (18 years or older) family members living with you?

(Perc. Freq.)

1. Under $5,000 (0.8)	6. $25,000-29,999 (8.3)	11. $60,000-69,999 (7.1)
2. $5,000-9,999 (1.0)	7. $30,000-34,999 (10.7)	12. $70,000-79,999 (4.8)
3. $10,000-14,999 (3.0)	8. $35,000-39,999 (10.7)	13. $80,000-89,999 (3.6)
4. $15,000-19,999 (3.9)	9. $40,000-49,999 (15.9)	14. $90,000-100,000 (2.7)
5. $20,000-24,999 (5.8)	10. $50,000-59,999 (13.1)	15. More than $100,000 (8.5)

n = 887

Median: @ $48,972.46

Q-18 About how many total hours per week do you and other adult members of your household spend working at income generating employment?

_________ Hours

Mode:	40.0 hrs./wk	n = 937
Mean:	59.60	
Std. Dev	45.98	

Q-19 What type of residence do you live in?

	Perc. Freq.	n = 806
1. House	84.6	
2. Apartment	4.3	
3. Condominium/Townhouse	8.1	
4. Studio	0.1	
5. Other ______________________	2.9	

Q-20 Do you own or rent your residence?

	Perc. Freq.	n = 921
1. Own	84.5	
2. Rent/lease	15.5	

Is there anything we may have overlooked? Please use the space below for any additional comments you would like to make concerning Denver's air quality and its influence in your life.

4.4 Descriptive Statistics for Questions Unique to VERSION E

Figure 4.5, which can be found at the end of this section, displays results for the questions unique to version E. The numbering of the questions in this facsimile correspond to the order in which they appeared in the actual survey. They should ***not*** be confused with the numbering of the questions in the facsimile of Figure 4.5. As discussed in Chapter 3, these questions were intended to provide a more complete context and to provide some assistance in fully defining the commodity to be valued.

Questions (Q-4) and (Q-5) in Figure 4.5 appeared at the end of Section I in version E. Question (Q-4) asked the respondents how much attention they pay to the visibility and odor of the air. Using a seven step scale, the average response was a rating of 3.03 with a standard deviation of 1.56. Question 5 asked if more or less money should be spent to reduce air pollution. Almost 70% of the respondents said that more should be spent to reduce air pollution and another 25.4% indicated that spending levels should remain the same.

Section II of this version contained additional questions designed to help define the market and the commodity. Question 8 asked for respondents opinions on what would happen to Denver's air pollution problem over the next 5-10 years if nothing further is done. Approximately 95% of the households surveyed said that the problem is likely to become worse. They were next queried as to the actions they had taken or intended to take in relation to air pollution reduction activities. The two most common actions households had taken with respect to pollution problems were reducing the use of fireplace and/or woodburning stoves and using oxygenated fuels (Q-9). Both are mandated programs in the City of Denver.

In Q-6, six possible sources of air pollution are listed. Using a 7 point scale, respondents rated automobiles as having the highest contribution to air pollution (mean=5.96) and fireplaces (mean=4.2) as having the lowest contribution. The most common other sources of pollution mentioned were diesel trucks and cars which, at present, have no required emission standards.

According to question 11, a respondents average time of residency in the Denver area was about 22 years. The modal category was equally divided between 30-40 years and 10-20 years. About 73% of the respondents to question 12 felt that the amount of attention given to air pollution problems in the Denver area was appropriate.

Section III of this version contained an alternative format for the elicitation of WTP and WTA values (as discussed in previous chapters and analyzed in the following chapter). It also contained extra information (a sample of which is provided in Figure 4.5) and one additional question. Before answering valuation questions, respondents were asked for their opinion on Denver's air pollution problems over the next 5-10 years if a stricter air pollution program were to be implemented. Seventy-eight percent of respondents indicated they felt Denver's air pollution problems would be improved if some type of tougher program were to be mandated.

The final section of this survey variant included three "demographic" questions in addition to those of the other variants. Question 19, tying in with the referendum voting format of the previous section, asked whether or not the respondent was actually registered to vote. Out of the 111 people answering the question 84.7% were registered. Question 20 asked if there was a "view" from the respondents' home of a mile or more. Approximately sixty-seven percent said they had some view, while 24.3% had some view and 8.4% had no view. Finally, we asked respondents to indicate the degree to which certain activities and interests applied to them. An interesting observation is that over 60% of the respondents indicated that they "definitely" spend a lot of time out of doors on weekends. In the category of "someone who trusted what experts say about science and technology", 41.4% circled 3 or lower on the seven point scale, where the lower bound was labelled as "Not At All."

Figure 4.5

Facsimile of Descriptive Statistics for Version E

(Perc. Freq. = Percent of Frequency of Occurrence)

I. THE ISSUES

Q-4 Generally speaking, how much attention do you pay to how far you can see and the odor of the air on a given day?

	A Lot 1	2	3	4	5	6	Not at All 7		
								Mean	3.03
(Perc. Freq)	(20.5)	(23.2)	(16.1)	(19.6)	(15.2)	(4.5)	(0.9)	Std. Dev.	1.56
								n = 112	

Q-5 Do you think more or less money should be spent trying to reduce air pollution in the Denver area?

	Perc. Freq.	
1. More money	(69.6)	n = 102
2. Same amount of money	(25.5)	
3. Less Money	(4.9)	

II. HOW DO YOU RATE DENVER'S AIR?

Q-8 In your opinion, if Denver does nothing further about air pollution and the Brown Cloud, over the next 5-10 years, the problem is likely to

	Perc. Freq.	
1. get much worse.	(68.5)	n = 111
2. get somewhat worse.	(26.1)	
3. stay about the same.	(5.4)	
4. get somewhat better.	(0.0)	
5. get much better.	(0.0)	

Q-9 Among the following actions, which have you and your household taken or intend to take? (Please circle all that apply.)

	Percent of Total Respondents
1. Reduced the use of your fireplace and/or wood-burning stove.	(69.4)
2. Taken some other form of transportation (i.e. bus, bicycle, walk, etc.) rather than your car to work or shopping.	(19.8)
3. Car-pooled.	(22.5)
4. Used oxygenated gasoline.	(79.3)
5. Did not drive one day per per week (to work or other places).	(27.9)
6. Other (please specify)__________________________	(6.3)
	n = 111

Q-10 To what degree do you think the following causes contribute to the air pollution problem in the Denver metro area?

Very Low Contribution (1) … Very High Contribution (7)

(Perc. Freq)

	1	2	3	4	5	6	7	Mean	Std.Dev.
1. Automobiles, n = 111	(0.9)	(0.0)	(7.2)	(9.9)	(10.8)	(18.0)	(53.2)	5.96	1.39
2. Fireplaces, n = 112	(3.6)	(10.7)	(16.1)	(25.9)	(23.2)	(14.3)	(6.3)	4.22	1.50
3. Power plants, n = 109	(0.9)	(8.3)	(9.2)	(13.8)	(28.4)	(15.6)	(23.9)	5.03	1.59
4. Industry, n = 112	(0)	(3.6)	(9.8)	(12.5)	(21.4)	(26.8)	(25.9)	5.36	1.43
5. Weather Conditions, n = 111	(5.4)	(5.4)	(9.0)	(13.5)	(19.8)	(18.9)	(27.9)	5.05	1.78
6. Other, n = 28 (Please specify)__________________________	(3.6)	(10.7)	(7.1)	(0)	(10.7)	(10.7)	(57.1)	5.64	1.99

Q-11 How long have you lived in the Denver metro area?

___________ YEARS ___________ MONTHS

Mode:	10-20, or 30-40 yrs.(an even split)	Mode:	4, or 6 mos.
Mean:	21.55 yrs.	Mean:	5.92 mos.
Std. Dev.:	15.30	Std. Dev.:	5.10

n = 111

Q-12 How do you feel about the amount of attention given to the air pollution problems of the Denver metro area?

	Too Little					Too Much		
	1	2	3	4	5	6	7	Mean: 3.44
(Perc. Freq)	(9.0)	(11.7)	(30.6)	(33.3)	(9.0)	(2.7)	(3.6)	Std. Dev. 1.35

n = 111

III. THE VALUE OF CLEAN AIR TO YOU

Future extensions of the Better Air Campaign could also include tighter emission standards on automobiles and tighter standards on industrial pollution. Catalytic converters or other forms of filtering smoke from wood burning stoves and fireplaces could become mandatory. Such actions could lead to higher prices for certain commodities (gasoline and/or wood-burning installations, for example).

Q-13 In your opinion, if Denver implements a tough program for air pollution and the Brown Cloud, over the next 5-10 years, the problem is likely to

	Perc. Freq.	
1. get much worse.	(4.6)	n = 109
2. get somewhat worse.	(4.6)	
3. stay about the same.	(12.8)	
4. get somewhat better.	(58.7)	
5. get much better.	(19.3)	

IV. ABOUT YOU

Q-19 Are you registered to vote?

	(Perc. Freq.)	
1. YES	(84.7)	n = 111
2. NO	(14.4)	

Q-20 Can you see a mile or more from your home or is it located in an area where you just see nearby structures?

	Perc. Freq.	
1. View	(66.7)	n = 111
2. Some View	(24.3)	
3. No View	(9.0)	

Q-21 Below is a list of phrases that describe different kinds of interests and activities that people have and do. As you read each one, please indicate the degree to which it applies to you.

	NOT AT ALL (Perc. Freq.)						DEFINITELY
1. "...someone who spends a lot of time out of doors on weekends."	1	2	3	4	5	6	7
n = 112	(0)	(0.9)	(6.3)	(13.4)	(18.8)	(28.6)	(32.1)
2. "...a birdwatcher."	1	2	3	4	5	6	7
n = 109	(35.8)	(22.0)	(13.8)	(8.3)	(12.8)	(5.5)	(1.8)
3. "...someone who trusts what experts say about science and technology."	1	2	3	4	5	6	7
n = 111	(7.2)	(10.8)	(23.4)	(24.3)	(17.1)	(10.8)	(6.3)
4. "...someone who is an environmentalist."	1	2	3	4	5	6	7
n = 111	(6.3)	(11.7)	(13.5)	(22.5)	(18.0)	(16.2)	(10.8)
5. "...someone who always votes in city elections."	1	2	3	4	5	6	7
n = 110	(18.2)	(10.0)	(7.3)	(14.5)	(15.5)	(12.7)	(21.8)

CHAPTER 5

Analysis of Hypothetical Bias and Context Effects

5.1 Introduction

There are two primary questions addressed in the analyses reported in this chapter. The foremost being: What is the extent of hypothetical bias when respondents lack prior market experience? The specific model of hypothetical bias proposed in Chapter 2 is analyzed under various econometric treatments, with suggestions as to which is the most appropriate for handling biases that can be associated with the CVM. The other addresses whether context, as implemented in the eight versions of the survey, affected the likelihood that a respondent would return the questionnaire, the likelihood that a respondent would provide a positive bid for improving air quality, and the magnitude of the response for those who did give meaningful responses. Additionally, we consider context effects within several of the survey versions; in particular, we address the question of whether respondents can meaningfully divide their overall values for improved air quality into separate components for visibility and health improvements.

5.2 Context Effect On Return Rate By Version

The middle column of Table 5.1 shows the questionnaire return rate for each version, adjusted for bad addresses. Version A is both the shortest and simplest version of the questionnaire so we used its return rate as the baseline against which to compare the other, more complicated versions. The return rates ranged between 67.9 and 74.4 percent and do not differ reliably from the baseline of 72.6 percent for Version A (χ^2 (6) = .42, ns.).

Table 5.1

Questionnaire Return Rates and Positive WTP Bid Rates by Version

Version	Percent Return	Percent Positive WTP Bids (# of Positive Bids, \$0 - \$+∞) Given Return
A	72.6	64.7
B	67.9	58.3
C	71.5	64.4
D	69.9	66.9
E	68.3	60.7

Table 5.1 (Con't)

Version	Percent Return	Given Return
F	71.2	71.0[1]
G	74.4	68.0
H	72.0	n.a.[2]

Versions B, D, and E are the only versions with response rates below 71.2 percent, however they are complicated and long relative to the other versions. Additional OAQPS health information text is contained in versions B and D, and of these two, version D asks three separate willingness-to-pay questions: visibility, health, and total. These separate components are also in Version C, and thus did not produce reliably lower response rates by themselves. Therefore it is the additional health information that may have produced a questionnaire that seemed unreasonably wordy and long to some respondents. Likewise, as outlined in Chapter 3, version E has an extensive number of additional context questions that made for a questionnaire that may have appeared too long for some households to fill out, and have the resulting effect of a lower response rate. In any case, there is no statistically significant difference between versions in rate of return due to context effects in this field study.

5.3 Context Effect On Positive WTP Bid Response Rate

An important goal of this project is to determine whether people can give meaningful monetary amounts for the value of improving air quality. Thus, an important question is whether the survey versions differ with respect to respondent likelihood of giving a positive WTP response. We define positive responses to be all positive willingness-to-pay values plus all zero values if the respondent gave a valid reason for a zero bid. Valid reasons included statements that “the brown cloud doesn't bother me” or “it's not worth anything to me to clean up the air.” Responses such as “industry or government should pay for it” indicated that cleaner air was valued by the respondent but that he or she didn't want to pay for it. Willingness-to-pay values

[1] There were two WTP questions for Version F. The reported meaningful WTP response rate is for the first question (reducing number of bad days). The meaningful WTP response rate for the second question (improving number of good days) was 55%, which is not significantly different than the rate for Version A.

[2] Version H contained only choice questions so respondents were not asked a WTP question.

of zero for those reasons appear to be protest statements rather than expressions of the respondents true value for improved air quality. Another indication that virtually all of the zero bids were protest bids or unthoughtful bids is that all but four of these respondents also did not state a finite amount they would be willing to accept to allow decreased air quality. It makes no sense to hold simultaneously that improved air quality has no value but that avoiding decreased air quality has infinite value. This common response pattern among those who stated a zero willingness-to-pay represents, we believe, strong cases of status quo bias. That is, people believe the present situation is the best of all possible situations.

The last column of Table 5.1 reports the percentage of respondents stating positive willingness-to-pay values given that they returned a questionnaire. For versions with separate questions valuing visibility and health effects (Versions B and D), positive bids to the total willingness-to-pay question were used to define meaningful responses. For Version F, which has separate willingness-to-pay questions for reducing the number of bad days and for increasing the number of good days, we used responses to the first question to define meaningfulness.[3] There are no significant differences across versions in the percentage of positive value responses, given that the respondent had returned a questionnaire. Thus, the different contexts represented in the different survey versions do not appear to influence the likelihood that the respondent will respond meaningfully. Respondents who wish to register a protest or who refuse to cooperate will do so no matter what survey version is used.

5.4 Analysis of Field Data

Although some respondents were asked for WTA as well as WTP values, only WTP was used for statistical analyses, because so few respondents gave finite WTA responses. Out of the 342 respondents who returned versions of the survey with the WTA question, the vast majority indicated that they would not accept any amount of money for a decrease in air quality (i.e., their WTA value was “infinite”). Only 28 (8.2%) respondents gave finite values for WTA.

---

[3] When positive responses to both questions are required for Version F, the percentage of meaningful responses drops to 55 percent. Everyone who gave a meaningful response to the second question also gave one to the first. The decrease in meaningful responses for the second question may simply be a fatigue effect. It may also be due to a belief that a response had already been given because the two questions appear at first glance to be very similar. Or some people may value increasing good days less than they value reducing bad days. Unfortunately, because order was confounded with type of question, these alternative hypotheses cannot be examined within these data.

Out of the 812 survey responses to versions A-G (the versions requesting WTP values), 283 (35%) of those respondents did not give positive WTP values. Three (1.06%) of these respondents indicated that they would not give a positive WTP value because air quality had no value for them. As mentioned earlier, the rest of the respondents either stated that they were not responsible for the problem or gave no reason. If more respondents had given zero dollar WTP bids for air quality, with the indication that the bid represented an actual zero value, then it would have been appropriate to perform a Tobit analysis to include these zero dollar bids in our regressions.

It is reasonable to conclude that, for many of the respondents, the refusal to bid indicated some other sentiment besides an actual zero value for air quality. Further, it is of interest to be able to predict when respondents are likely to give a positive WTP value, as well as to be able to include this 35% of the sample in the calculation of predicted values. A Probit analysis was performed to accomplish two goals: to identify variables that predict whether respondents will bid a positive WTP amount, and to provide a selection bias variable that would allow the inclusion of missing respondents in the model to predict WTP values.

Survey Variables Used in the Analyses

Three types of variables were used to model the WTP values: (a) context design variables, (b) sociodemographic variables, and (c) air quality rating variables. Each type is described in turn.

Context Design Variables Dummy codes were used to indicate the design features of each survey version. The variable **HEALTH** denotes whether or not the survey version contained the additional health information supplied by OAQPS; Versions B and D contained this information. The value of 1 for variable **THREE** indicates that the survey version contained three separate WTP questions: one for visibility, one for health effects, and one for total air quality; Versions C and D each have the three questions. The variable **VOTING** denotes Version E which framed the WTP question in terms of a referendum. The variable **FREQ** indicates Version F which displayed the improvement in air quality in terms of changes in frequency distributions. Finally, the variable **COMMOD** indicates Version G which asked the WTP question for air quality after asking similar questions for two other commodities — a camera and a television.

Sociodemographics Standard sociodemographic variables included **AGE, GENDER** (1 = male, 2 = female), **EDUC**ation (see facsimile survey for categories), and **INCOME** (see facsimile survey for categories). In past research on responses to risks, an important variable is whether or not there are children living in the household. The variable **KIDS** indicates whether or not there are any children 18 years or younger living in the household.

Air Quality Ratings Several questions asked respondents to evaluate the typical air quality they experienced and how much the visibility and health problems bothered them. It is reasonable to ask whether people who were more bothered by air quality problems or who thought typical air quality was worse were willing to pay larger amounts for improved air quality. **VISRT** is the rating on a seven-point scale of "how bothered have you been by the Brown Cloud's effect on what you can see in the distance (mountains, buildings, etc.)?" **HEALTHRT** is the rating on a seven-point scale of "have you or your family been bothered by any health problems which you believe to be caused or aggravated by Denver's air pollution?" In both cases higher ratings mean the respondent was more bothered.

Finally, **AQRATING** is the respondent's rating of the "current overall air quality (thinking of both health and visibility) on an average winter day" on the air quality ladder defined by the photographs enclosed with the survey.

Probit Results

Table 5.2 shows the results of the Probit analysis.[4] AGE, INCOME, EDUC and KIDS all identified respondents who were likely to give a positive WTP value. Old age, low income, low education level, and having many children all decrease the chances that a respondent will give a positive WTP value for air quality. Note that neither concern for visibility VISRT or health HEALTHRT significantly increased the probability of a positive WTP. These variables would be expected to have a positive effect if refusal bids were indicative of zero WTP values. Rather, this result suggests that these respondents do not accept responsibility for the problem. Note that if either of these concern variables are included in the positive WTP regression reported below they show a positive and significant correlation with WTP.

WTP Distribution Figure 5.1 shows the distribution of refusal (zero) and WTP values, pooled across versions. The WTP value distribution is positively skewed, with a very long and thick positive tail, meaning that some respondents indicated that they were willing to pay quite large amounts for an improvement in air quality. When untransformed WTP values are regressed on the context and demographic variables, (see Table 5.3 for results of all WTP regressions), the resulting error distribution is also quite positively skewed (Figure 5.2, first distribution).

The WTP distribution appears to have a log normal shape, indicating that a log transformation of the WTP data would result in a more normal error distribution. Indeed, the regression of log (WTP) on the independent variables results in a more normal error distribution

---

[4]The three respondents who gave "actual" zero dollar values for WTP were not included in the Probit analysis.

Table 5.2

Probit Analysis

(Parameters in bold are significant at $p<.05$)

Variable	Parameter Estimate	t
INTERCEPT	.214	.494
THREE	-.072	-.665
VOTING	.130	1.241
HEALTH	-.000	-.001
COMMOD	.244	1.632
FREQ	.017	.118
VISRT	-.001	-.011
HEALTHRT	.037	.894
AVERT	-.041	-1.349
AGE	**-.010**	**-2.471**
GENDER	.085	.825
INCOME	**.070**	**2.095**
EDUC	**.049**	**3.031**
KIDS	**-.17**	**3.268**

Figure 5.1

WTP Frequencies (In Dollars)

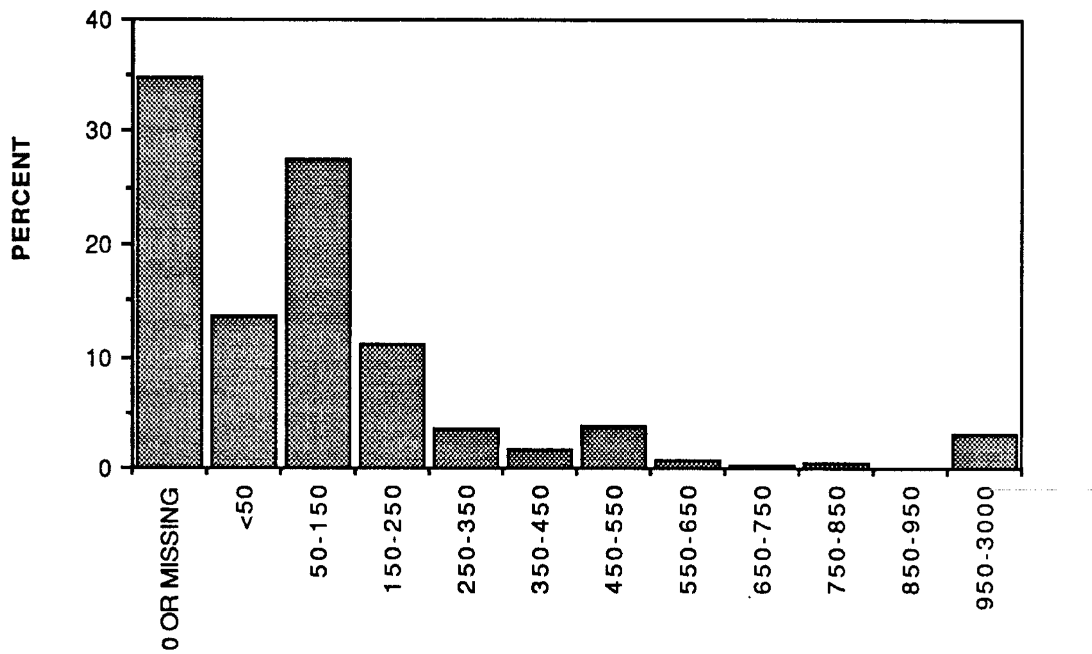
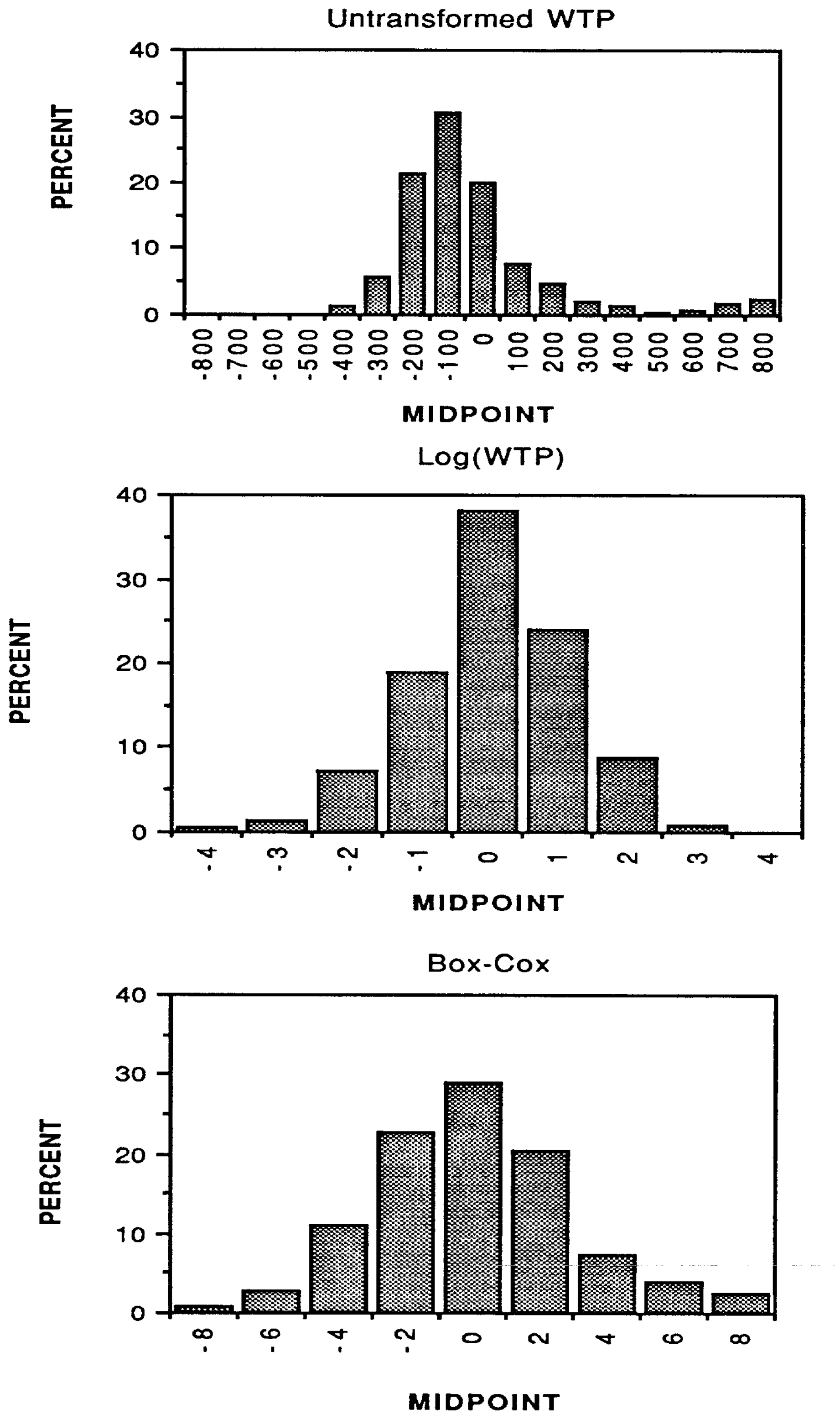


Figure 5.2

Residuals for WTP Regressions



(Figure 5.2, second distribution), although this distribution exhibits a slight negative skew. In order to make this log (WTP) error distribution normal, some of the lower WTP values would have to be trimmed. A more satisfying solution is to transform the data using a Box-Cox procedure, which transforms the data using the best fitting α, such that Box-Cox = $(B^{\alpha}-1)/\alpha$.

When the Box-Cox WTP values are regressed on the independent variables, the resulting error distribution is symmetrical (Figure 5.2, bottom distribution). The estimated α is .2135. An α of 1 produces a linear transformation of the data; an α of 0 results in logarithmic transformation.

Regression Results

After the Probit analysis, three regressions were run, all of which included the selection bias variable (λ) from the Probit analysis. The first regression used untransformed WTP values, the second used log transformed WTP values, and the third used the Box-Cox transformation. The parameter estimates from these regressions are listed in Table 5.3. The mean predicted values from these three regressions are presented in Table 5.4. Table 5.4's reported means demonstrate that the mean for the model using untransformed WTP is higher than the means using transformed WTP. The mean for the log(WTP) model is somewhat lower than the mean for the Box-Cox model. Considering the distribution of the residuals for these two models, this should come as no surprise. Predicted values were also computed for those respondents not giving a positive WTP value (the "nonbidders"); for the untransformed WTP regression, this mean predicted value is higher for the nonbidders than for bidders. For the log(WTP) and Box-Cox regressions, the predicted means are lower for nonbidders than for bidders.

In all of the regressions, the selection bias, l, variable is not a significant predictor of WTP values. Thus, in this case excluding bid refusals from the data analysis would not bias the estimated WTP equation.

Of the survey design variables, FREQ is the only reliable predictor of WTP values (at the 5% level), with COMMOD and THREE only weakly significant (at the 10% level) in the Box-Cox regression. Version F WTP values, used in the analysis of FREQ, were computed by adding two independent half-step values that respondents gave for two frequency distribution shifts. This effect could result from diminishing marginal valuation of air quality improvements, but more likely is the result of superadditivity.

Note that the upward bias in version C and D associated with THREE, where respondents first provided separate health and visibility values before giving a total WTP, also indicates a super additivity problem. Version G, in which respondents valued two commodity trades before valuing air quality, also resulted in marginally higher air quality WTP values. Respondents here

Table 5.3

Parameter Estimates and t's by Variable for Models of WTP

(Bold entries are significant at p < .05)

Variable	Untransformed	Log	Box-Cox
INTERCEPT	-134.374	**3.686**	**4.732**
	(-1.061)	**(5.399)**	**(3.114)**
HEALTH	-21.092	-.226	-.410
	(-.630)	(-1.254)	(-1.021)
THREE	**61.349**	.268	.651
	(1.807)	(1.465)	(1.599)
VOTING	34.302	-.115	-.239
	(.788)	(-.492)	(-.458)
FREQ	**84.691**	**.421**	**1.160**
	(1.991)	**(1.838)**	**(2.273)**
COMMOD	57.748	.216	.798
	(1.424)	(.990)	(1.641)
AGE	-.210	-.003	-.001
	(-.149)	(-.464)	(-.079)
GENDER	2.854	.015	.117
	(.101)	(.100)	(.348)
EDUC	**20.135**	.067	**.211**
	(1.997)	(1.235)	**(1.743)**
INCOME	**23.311**	**.104**	**.286**
	(4.077)	**(3.373)**	**(4.179)**
KIDS	-19.191	-.047	-.069
	(-.744)	(-.342)	(-.223)
AVERT	**-36.104**	-.119	-.333
	(-2.052)	(-1.258)	(-1.580)
LAMBDA	94.871	-.423	-.461
	(.710)	(-.588)	(-.287)
	R^2 = .09	R^2=.11	R^2=.14

Table 5.4

Mean Model Estimates of WTP

(Nonbidders did not give a positive WTP bid)

	Untransformed	Log	Box-Cox
<u>Separate Means</u>			
Nonbidders	244.22	52.68	82.77
Bidders	219.29	94.71	119.72
<u>Grand Mean</u>			
Nonbidders' bids counted as $0	146.30	63.18	79.87
Using Nonbidders' Pred Values	231.76	73.70	101.24

may have anchored on the value of the commodity trades, feeling "morally" obliged to put more value on air quality.

INCOME and EDUCATION are the only demographic variables that are found to be able to help significantly predict WTP values, with higher income and higher educational level predicting higher WTP values. (One exception is EDUC in the log(WTP) regression, which is below significance at the .05 level.) Respondents' ratings of a typical winter day (AQRATING) are predictive of WTP values in the linear regression only, suggesting that this rating is in fact not a reliable predictor of WTP values.

Commodity/Air Quality Comparisons In versions G and H, respondents were asked to evaluate trades in air quality, television sets, and cameras. Respondents to version G were asked to give WTP values for the three trades. Their mean WTP values for the camera trade were lower, overall, than their mean WTP values for the T.V. trade, and air quality WTP values were higher, overall, than either of the commodity WTP values.

To compare responses on the two survey versions, it is necessary to convert the Version G monetary values into imputed choices that are comparable to the Version H choice responses. The imputed choices are obtained by comparing the WTP values for the commodity trades and the air quality trades. If the respondents gave a higher value for the air quality trade than for the commodity trade, they are given a "1" (prefers air quality trade to camera/T.V. trade), if they gave the same value for the air quality and commodity trade, they are given a "2" (value the trades equally), and if they gave a higher WTP value for the commodity trade than for the air quality trade they are given a "3" (prefers commodity trade to air quality trade) for that trade.

Figure 5.3 shows the percentage of respondents in the choice categories, for the camera/air quality trade comparison and Figure 5.4 shows the comparable percentages for the T.V./air quality trade comparison. Although most respondents favor the air quality trade over the commodity trades, there is a significant difference in choices due to version type ($\chi^2(1) = 14.3$, $p < .0001$ for the camera/air quality comparison and $\chi^2(1) = 23$, $p < .0001$ for the T.V./air quality comparison). The choice differences due to reward type indicate a higher relative valuation of the commodity trades, and a lower relative valuation of the air quality trade, when respondents receive Version G of the questionnaire. Specifically, respondents in Version G are relatively more likely to value the camera/T.V. trade more than the air quality trade, and respondents in version H are more likely to value the air quality trade above the T.V. trade.

Figure 5.3
Camera Versus Air Quality Ratings
By Version

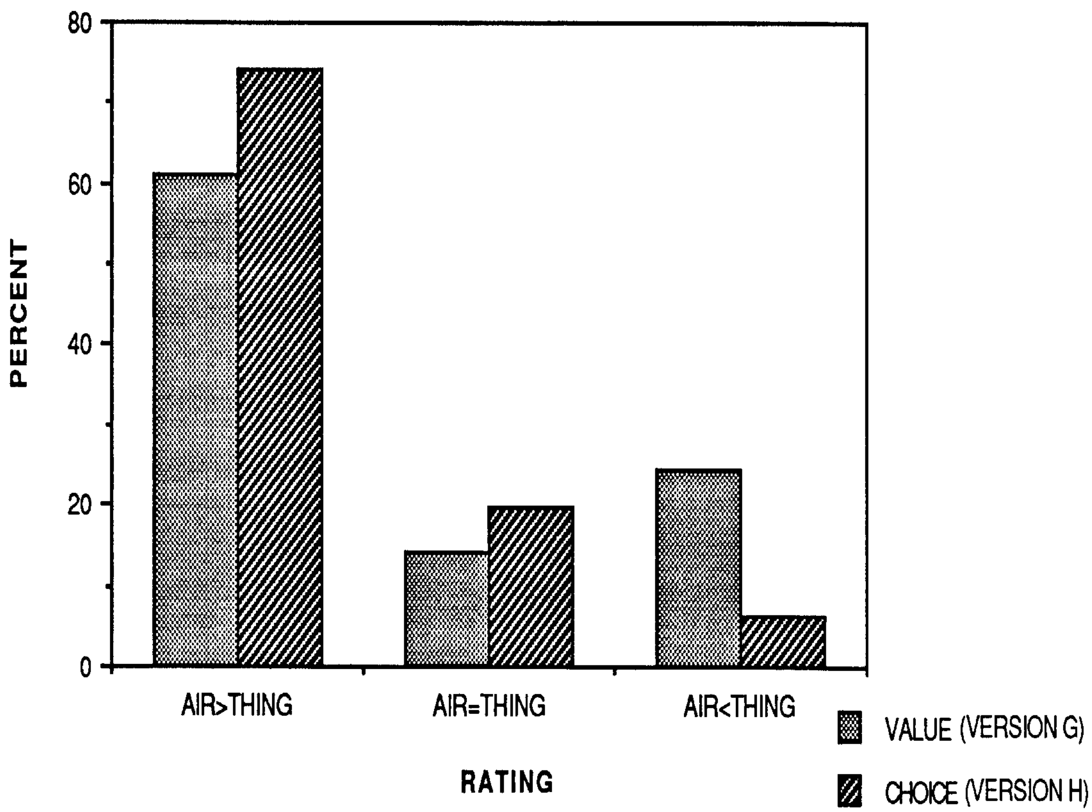
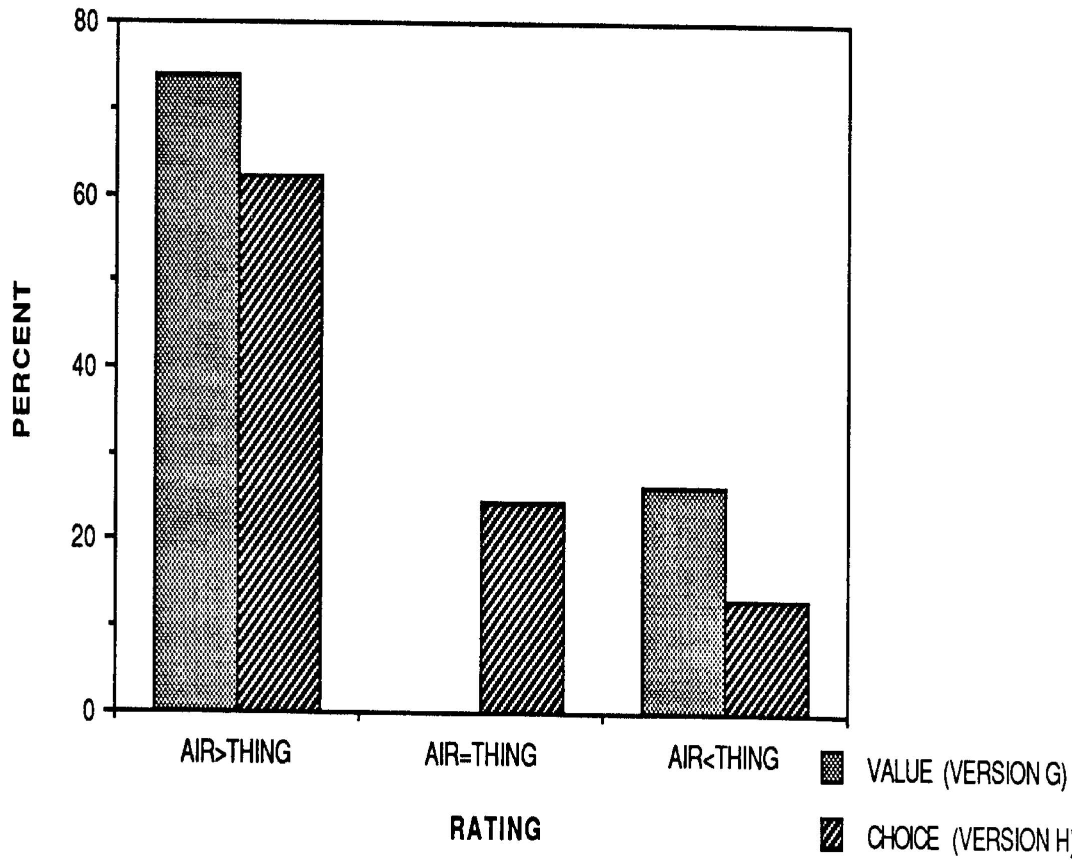


Figure 5.4
T.V. Versus Air Quality Ratings
By Version



When the analyses are run using just ratings 1 and 3 (i.e., no ties), the version/rating difference is significant for the camera ratings ($\chi^2(1) = 13.22$, $p < .001$) but not for the T.V. ratings.

In summary, the comparison of actual and imputed choices from Versions G and H indicates that respondents, on average, preferred the air quality trade to either commodity trade, but the probability that they would prefer one trade to another depended on version type. When respondents were asked to provide WTP values, with no explicit instruction to compare values across the three trades (Version G), they were more likely to value the commodity trade over the air quality trade and less likely (for the T.V trade, at least) to value the air quality trade over the commodity trade.

Ratings For the "how bothered are you..." questions (VISRT and HEALTHRT), the average rating on a seven-point scale for visibility, 5.4, is reliably greater than the average rating for health, 3.3, (*t* (377) = 30.9, p < .0001). The correlation between the two ratings is statistically reliable (r = .37, p < .0001). Table 5.5 shows the mean ratings, the difference, and the test statistic for the difference for each of the six photographs.

All respondents also rated six photographs on an air quality ladder separately in terms of visibility and health. For all the photos, the difference between ratings for visibility and health are generally small, with mean differences ranging from -.08 to .25. The overall visibility-health rating differences is nevertheless statistically significant (*t*(827) = 7.68, p < .0001).

Table 5.5

Ratings of Photographs on Air Quality Ladder

Photograph	Vis	Health	Vis-Health	t	r
A	3.1	2.9	0.19	5.88	.60
B	5.6	5.4	0.18	4.99	.65
C	1.3	1.4	-0.08	-1.57	.41
D	6.6	6.3	0.25	7.34	.51
E	2.1	2.0	0.09	2.79	.55
F	3.3	3.1	0.16	4.77	.67

Although the mean ratings for visibility and health are not substantially different they are statistically different because of very high statistical power afforded by the large sample size (about 780 for each mean) and the within-subject design. The correlations between the visibility and health ratings are also reported for each photograph in Table 5.5. With one exception (Photo C) they are all much larger than the correlation between the bothersome ratings of visibility and health reported above. Hence, the general impression is that respondents are ***not*** able to make fine discriminations between visibility and health effects when evaluating photographs of air conditions.

WTP Values within Survey Version Several of the survey versions asked respondents who gave WTP values to give the percentage of that value attributed to visibility and health. The mean share for visibility (27.2%) was reliably lower than the mean share for health (48.3%, *t*(229) = -8.53, p < .0001). 92% of respondents who gave positive numbers for both values gave percentages that added up to 100; the majority of the remaining respondents gave percentages that added up to less than 100%. 23% of respondents giving values for both percentages gave values of “0” for both. Thus, respondents attributed 27.2% of their total value to visibility 48.3% to health, and 24.5% to other sources of value. If we use the Box-Cox predicted values, including those for non-bidders, then the mean total bid of $101.24 would be distributed as follows: $27.54 to visibility, $48.90 to health, and $24.80 to other. Note that this approach implies a significant difference between health and visibility values and avoids the superadditivity problem discussed below.

It is interesting to note that these same respondents indicated they were more bothered by visibility effects than by health effects (5.4 versus 3.3). Thus, respondents are more bothered by the visibility effects of the brown cloud but more of what they are willing to pay for cleaner air is attributable to a desire to reduce health effects. There are several ways to interpret this result. One possibility is that the responses are illogical. Another is that even though respondents have not themselves been bothered by health effects they believe it is important to pay for eliminating health effects that others may be experiencing or that they may experience in the future. Some people may believe it is morally more important to reduce health problems than visibility problems. Since percentages attributed to visibility and health do not add to 100%. Respondents may simply have been careless or the remainder (about 8%) may be attributable to other concerns about air pollution.

Two of the survey versions (B & D) asked respondents to state separate WTP values for visibility and health for a one-step improvements on the air quality ladder. Although the mean for separate visibility values was lower than for separate health values, the difference was not statistically reliable ($t(191) = .78$). The sum of respondents’ visibility and health WTP values significantly exceeded their total WTP values for air quality ($t(190) = 2.51$, $p < .01$). Thus the superadditivity hypothesis is strongly supported in our results. Note, however, that the total WTP as given by respondents was only marginally significant and greater than in other survey versions.

The final WTP analysis compared respondents’ WTP values for the two frequency distributions presented in version F. One frequency distribution depicted a half-step improvement in air quality involving a reduction in the number of bad air quality days; the other distribution depicted a half-step improvement in the number of good air quality days. The WTP values for the reduction in bad days were significantly higher than the values for increasing the number of good days (t (104) = 2.08, $p < .04$).

5.5 Conclusions

The first goal of this study was to attempt to apply an understanding of hypothetical bias arising from lack of market experience obtained from laboratory experiments to a field application of the CVM. Mean estimated values using the Box-Cox transform which attempts to eliminate hypothetical bias appear to be substantially smaller than those obtained by using a linear regression approach. In the case of air quality in Denver, selection bias resulting from bid refusals was not important. The treatment of refusal bids, assuming refusal bids are zeros or using the predicted values for non-bidders, was of substantially lesser importance than correcting for the skewed error distribution.

A second and related goal of this study was to provide information about the effect of contextual variables on values elicited by CVM questions about air quality. Our overall conclusion is that, contrary to the expectations of theorists such as Fischhoff and Furby (1989), for the most part, contextual variables in our survey did not greatly affect respondents' WTP values, implying that values are crystallized in this particular case. None of the observed context effects changed the order of magnitude of air quality values.

Respondents to the version with air quality frequency distributions did give significantly higher WTP values. In this case respondents evaluated two frequency distributions, each representing a half-step change in air quality, so that the sum of their two values would theoretically equal a WTP value for a one-step increase in air quality**.** However, since this was an attempt to split out the value of reducing the number of poor days from the value of increasing the number of good days, it is not surprising that the sum of the values exceeded that obtained in other versions given the problem of superadditivity. Note also that the sum of separately obtained health and visibility values exceeded the total value obtained from the *same* respondents. Thus superadditivity appears to be a consistent problem for respondents that requires great care in the design and interpretation of surveys attempting to obtain values. Use of percent splits rather than use of separate values to obtain component values may be an appropriate design strategy.

REFERENCES

Box, G.E.P., and D.R. Cox, (part 2), 1964. "An Analysis of Transformations," *Journal of the Royal Statistical Society*, Series B, XXVI, 211-243.

Bishop, R.C., and T.A. Heberlein, 1978. "Measuring Values of Extra-Market Goods: Are Indirect Measures Biased?" *American Journal of Agricultural Economics*, Vol. 61, No. 5, 926-930.

Brookshire, D., M. Thayer, W.D. Schulze, R. d'Arge, 1982. "Valuing Public Goods: A Comparison of the Survey and Hedonic Approaches," *American Economic Review*, Vol. 72, No. 1., 165-177.

Cox, J.C., V.L. Smith, and J.M. Walker, 1989. "Reward Saliency, Dominance, and Decision Costs in Experimental Economics." Paper presented at Economic Science Association, Fall Meetings.

Coursey, D., J. Hovis, and W.D. Schulze, 1987. "On the Supposed Disparity Between Willingness to Accept and Willingness to Pay Measures of Value," *Quarterly Journal of Economics*, 679-690.

Cummings, R.G., D.S. Brookshire and W.D. Schulze, 1986. *Valuing Environmental Goods: An Assessment of the Contingent Valuation Method*, Rowman & Allanheld, publishers.

Dickie, M., A. Fisher, and S. Gerking, 1987. "Market Transactions and Hypothetical Demand Data: A Comparative Study," *Journal of the American Statistical Association*, Vol. 82, No. 397, 69-75.

Dillman, D.A., 1978. *Mail and Telephone Surveys: The Total Design Method*, John Wiley & Sons, N.Y.

Desvousges, W.H., V.K. Smith, and A. Fisher, 1987. "Option Price Estimates for Water Quality Improvements: A Contingent Valuation Study for the Monongahela River," *Journal of Environmental Economics and Management*, Vol. 1 14:3,148-267.

Doane, M., G.H. McClelland, W.D. Schulze, C.K. Woo, 1989. "The Value of Electrical Power Outages: A Contingent Valuation Study," Report submitted to Niagara Mohawk Electric Power Company.

Fischhoff, B. and L. Furby, 1988. "Measuring Values: A Conceptual Framework for Interpreting Transactions with Special Reference to Contingent Valuation of Visibility," *Journal of Risk and Uncertainty*, forthcoming.

Freeman, M.A. III, 1987. "Non-Use Values in Natural Resource Damage Assessment," Manuscript, Bowdoin College.

Greenley, D.A., R.G. Walsh, and R.A. Young, 1981. "Option Value: Empirical Evidence from a Case Study of Recreation and Water Quality," *Quarterly Journal of Economics*, Vol. 96. No. 4, 657-672.

Hanemann, M.W., 1989. "Willingness to Pay and Willingness to Accept: How Much Can They Differ?" Manuscript, University of California, Berkeley, California.

Heckman, J., 1979. "Sample Selection Bias as a Specification Error," *Econometrica.*

Hogarth, R.M, ed., 1982. *Question Framing and Response Consistency*. Jossey-Bass, Inc., Publishers. San Francisco, California.

Irwin, J.R., G. H, McClelland and W.D. Schulze, 1989. "Hypothetical and Real Consequences to Low Probability Risk," Submitted to *Journal of Behavioral Decisionmaking*.

Lichtenstein, S., and P. Slovic, 1971. "Reversals in Preference Between Bids and Choices in Gambling Decisions," *Journal of Experimental Psychology*, 98, 46-55.

Lichtenstein, S., and P. Slovic, 1973. "Response-induced Reversals in Preference in Gambling: An Extended Replication in Las Vegas," *Journal of Experimental Psychology*, 101, 16-20.

Maddala, G.S., 1983. *Limited-Dependent and Qualitative Variables in Econometrics*. Econometric Society Monographs, Cambridge University Press, Cambridge, Mass.

Mitchell, R.B. and R.T. Carson, 1988. *Using Surveys to Value Public Goods The Contingent Valuation Method*. Published by Resources for the Future, Washington, D.C.

Noelle-Neuman, E., 1970. "Wanted: Rules for Wording Structural Questionnaires," *Public Opinion Quarterly*, 34, 191-201.

Schulze, W.D., et al., 1983. "The Economic Benefits of Preserving Visibility in the National Parklands of the Southwest," *Natural Resources Journal*, Vol. 23, 151-173.

Schulze, W.D., 1988. "Use of Direct Methods for Valuing Natural Resource Damages," Paper prepared for the Conference on Natural Resource Damage Assessment, June 15-17, 1988.

Schuman, H. and S. Presser, 1981. *Questions and Answers in Attitude Surveys: Experiments on Question Form, Working and Context*. Academic Press, Inc. Harcourt Brace Jovanovich, Publishers.

Smith, V.K., W.H. Desvousges, and A. Fisher, 1986. "A Comparison of Direct and Indirect Methods of Estimating Environmental Benefits," *American Journal of Agricultural Economics*, Vol. 68:2

Smith, V.K., and W.H. Desvousges, 1987. "An Empirical Analysis of the Economic Value of Risk Changes," *Journal of Political Economy*, Vol. 95, No. 1, 89-114.

Stewart, T. R., P. Middleton, and D. Ely, 1983. "Judgments of Visual Air Quality: Reliability and Validity. *Journal of Environmental Psychology*, 3, 129-145.

Stewart, T.R., P. Middleton, P., M. Downton, and D. Ely, 1984. "Judgments of Photographs vs. Field Observations in Studies of Perception and Judgment of the Visual Environment." *Journal of Environmental Psychology*, 4, 283-302.

Tversky, A. and D. Kahneman, 1974. "Judgment Under Uncertainty: Heuristics and Biases," *Science*, 185, 1124-1131.

Tversky, A., and D. Kahneman, 1981. "The Framing of Decisions and the Psychology of Choice," *Science*, 211, 453-458.