A39

91-51.4

The Grand Canyon Visibility Benefits Study

William E. Balson
Decision Focus, Inc.
Los Altos, California

Richard T. Carson
University of California, San Diego
La Jolla, California

Robert C. Mitchell
Clark University
Worcester, Massachusetts

AIR & WASTE MANAGEMENT
ASSOCIATION
SINCE 1907

For Presentation at the
84th Annual Meeting & Exhibition
Vancouver, British Columbia
June 16 - 21, 1991

INTRODUCTION

The general context for this paper is a major policy controversy over the steps to be taken to limit visibility impairment at national parks. The particular context is the debate over whether the Navajo Generating Station (NGS) should install scrubbers to lessen its contribution to visibility impairment at the Grand Canyon National Park. The U.S. Environmental Protection Agency (EPA) has recently proposed a Best Available Retrofit Technology (BART) action which would require the NGS to cut its emissions by 70-90%, depending on the final rule. Action of this magnitude would require the NGS to spend more than a billion dollars to install scrubbers. This issue pits the EPA, the National Park Service (NPS), and the Environmental Defense Fund against the various owners of the NGS, namely, the Bureau of Reclamation and several large public utilities, including the Salt River Project (Phoenix) and the Los Angeles Department of Water and Power.

This paper describes our work, the Grand Canyon Visibility Benefits (GCVB) Study[1], to value visibility changes of the type which might result from EPA's proposed BART action. Our work was sponsored by the Salt River Project.* EPA and NPS have also sponsored a contingent valuation survey by Chestnut and Rowe (hereafter CR Survey)[2] to value visibility benefits in the Southwestern Parks and an effort by Rowe, Chestnut, and Skumanich (hereafter RCS Report)[3] to extrapolate the results of that survey to obtain a benefit estimate for Grand Canyon visibility improvements which might result from an NGS BART action. Much of the design of the GCVB study was motivated by what we considered flaws or weaknesses in earlier drafts of the CR Survey and RCS Report. As a result, this paper frequently compares and contrasts the design of the GCVB study with that of the CR Survey and the RCS Report. We believe that this approach will help clarify many of the key issues surrounding the valuation of visibility improvements in national parks.

MEASURING VISIBILITY BENEFITS

Contingent Valuation

Environmental amenities like air and water quality can not be bought and sold in the open market. Economists have developed several techniques for measuring the value of these so-called non-marketed goods. When the primary value of the non-marketed good comes from its direct use by the public, a number of different techniques can be used. When there are thought to be substantial nonuse values, then contingent valuation (Mitchell and Carson, 1989)[4] is the only economic valuation technique capable of correctly, from the perspective of economic theory, measuring both use and nonuse benefits of visibility improvements. Contingent valuation (CV) is a survey based approach which sets up a market in which a respondent can make purchase decisions involving the environmental amenity of interest. The respondent is queried about his or her willingness to pay (WTP) for this environmental amenity. The choice is often placed in the context of how the respondent would vote if the cost to them of providing the amenity was $x. Both the CR Survey and the GCVB Study use the contingent valuation method. The method is increasingly accepted by economists and the courts (Kopp, Portney, and Smith, 1990)[5] as a valid way to measure the public benefits of natural resources.

---

*The opinions expressed in this paper are solely the responsibility of the authors and should not necessarily be taken to be those of the Salt River Project. A number of individuals and organizations assisted in various phases of the research reported here, including Michael Conaway, Michael Hanemann, Annette Hulse, Ray Kopp, Kerry Martin, John Molenar, Stanley Presser, Paul Ruud, and Greg Witherspoon. They too should remain blameless for any errors we have made.

Design Choices

As with any research method, the validity of any individual CV study depends on the degree to which the potential sources of bias and error are controlled. In the case of contingent valuation, careful attention must be given both to economic theory issues and to a number of survey research issues. These issues include the wording of the questionnaire, the size, the design, and the selection of the sample, the execution of the survey, including the method of administration, and the analysis of the data. The wording, which involves the choice of scenario elements, the sequence of their introduction, and the language used to convey this framework to the respondent and to pose the valuation questions, is of obvious importance because it will influence the respondents' answers, including their willingness-to-pay amounts. The wording choices therefore require justification and, for studies used in policy purposes, must survive scrutiny by professionals and policy-makers.

In the course of designing a contingent valuation survey suitable for use in benefit/cost decision-making, the researcher inevitably must make and justify a number of design decisions which often have no obviously correct answers. Many of these decisions involve the scenario. The scenario is the market framework within which the respondent is asked to value the good. It includes such things as the payment vehicle, the duration of the payments, the nature of the good and how it will be provided, the method used to elicit the willingness-to-pay (WTP) amount, such as a payment card or an open-ended question, and conditions under which the good will be provided, such as the time of the year. The scenario used in a contingent valuation survey should be consistent with applied welfare theory and the nature of the good being valued. It must also be plausible and understandable to the ordinary respondent for whom the experience of stating a dollar value for a non-marketed good may be a novel experience.

Previous Work

The RCS Report estimated the economic benefits from winter visibility improvements in the Grand Canyon National Park that might result from a BART action requiring air pollution controls on the NGS. That report based its benefit estimates on the earlier CR Survey. The benefit estimates presented in the RCS Report are flawed because they are inappropriately derived from the CR Survey through questionable analytical assumptions and appear to substantially overstate the benefits of a possible BART action.

The CR Survey questionnaire was mailed to 3,345 households selected in five states: Arizona, California, Missouri, New York, and Virginia. A total of 1,647 responses were received. An insert contained 3 x 5 inch color photographs of four levels of visibility conditions during the summer at one or more of three national parks. The four levels were the reference level, one decreased visibility level, and two improved levels. The good was described as changes in average annual visibility range. Photographs for the reference level showing a range of 135 miles were described as illustrating the current annual range. Respondents valued large improvements to 200 and 250 miles and the prevention of deterioration to 115 miles. The study used a higher prices and taxes payment vehicle, a payment card elicitation method, and an annual payment in perpetuity.

Three versions of the instrument obtained WTP amounts for improving visibility in all the Southwestern national parks. Willingness to pay for the Grand Canyon was obtained by asking respondents to say what percentage of these WTP amounts they would allocate to the Grand Canyon. Although these three versions differ in the information provided to the respondents, Chestnut and Rowe say they found no statistically significant differences in responses for the three versions and therefore combined the results for these treatments, using those results in the RCS Report to EPA which estimates

Grand Canyon visibility benefits. The Grand Canyon estimates are based on approximately 700 usable cases from the three different treatments.

Two major problems with the CR Survey bear mention here. That study does not at any point specifically value the type of visibility change envisioned as a result of a Navajo BART action, i.e., a decrease in winter event days. Rather, the CR Survey values the aggregate of large changes in average annual visibility levels at all Southwest parks with the result that statistical techniques and a long chain of unverifiable assumptions must be used to impute a value for the visibility change that EPA projects will occur in the Grand Canyon due to a Navajo BART action. The primary source of proposed visibility change, a coal-fired power plant, is never mentioned to respondents. Furthermore, region-wide improvements -- the Southwestern Parks -- were the primary focus of the CR Survey, not the Grand Canyon. Respondents were asked for a single "percentage" to allocate their WTP amounts to Grand Canyon visibility.

A second problem concerns what respondents in the CR Survey were actually valuing. The intent of that study was to value visibility; the issue is whether their respondents were also valuing other presumed effects of air pollution such as health effects, harm to wildlife, harm to plants, and harm to geologic formations, *in addition to* visibility. The problem of separating visibility benefits from other types of benefits is driven by consideration of provisions of the Clean Air Act and has long been recognized as a difficulty by contingent valuation researchers working in the area of air pollution. Two approaches have evolved for dealing with this separability issue. One approach is to take an aggressive tact to informing respondents that they are only buying visibility improvements and that the health of plants, animals, and humans is not at risk or affected by the policy in question. The second is to take a statistical approach to separating the values of different types of benefits using an explicit hedonic pricing equation (Carson, Mitchell, and Ruud, 1990)[6]. The CR Survey did neither of these, and a substantial part of the benefits they attribute to visibility may in fact come from the respondents' wrongly believing that they will get other types of benefits if the visibility is improved.*

A number of other survey-related problems weaken the CR Survey as a basis for a NGS BART decision. These include:

- The results were extrapolated to the entire U.S. population despite the fact that the CR Survey was administered in only five states and was based on an inadequate sampling frame within those states.

- The study does not adequately address the problems in using a mail survey to administer a visibility contingent valuation study. One is the high level of illiteracy and semi-literacy in the United States (Mitchell and Carson, 1989)[4]; 10 to 20 percent of the population would find it hard or impossible to understand a questionnaire as complex as the CR instrument. Another is the possibility of selection bias inherent in mail surveys where those who look at it and choose to answer it may be more interested in the subject

* The CR Survey instrument told their respondents to give only their willingness to pay for visibility improvements. However, in a follow-up question (Q17), only 32% of their respondents indicated that their WTP amounts were "basically for the stated changes in visibility at national parks". Even this 32% is in doubt though, as 115 out of 449 respondents who gave this response did not follow instructions and later indicated that the average percentage of their WTP amounts which should be attributed to visibility was slightly less than 50%. Further, the CR Survey only made an attempt to separate visibility from a general "help other needs at the national parks" and did not make an attempt to separate "visibility" from several other phenomena such as acidic clouds damaging rock formations which our focus group work suggests many people initially see as closely intertwined with visibility degradation.

matter and have higher values for the good it describes than those who examine the questionnaire and choose not to answer it. A third is the lack of control over who fills the questionnaire out and how it is administered. A self-administered questionnaire may or may not be given the amount of serious attention necessary to arrive at a meaningful value for the good.

THE GCVB STUDY

By building on the knowledge gained from previous CV studies of visibility improvements, including our own work on Eastern air visibility (Carson, Mitchell, and Ruud, 1990)[6] and the CR Survey, a draft instrument was developed which was then tested and revised repeatedly until threats to its validity had been identified and satisfactorily addressed. In the course of the research reported here, we conducted focus groups, two telephone surveys, two survey pretests, and a pilot study of 202 persons. This series of research activities was designed to assess whether our initial critique of the CR Survey was correct and to help us understand the problems that would have to be overcome in designing a valid visibility contingent valuation study for the Navajo BART case (Balson, 1990)[7]. The focus groups explored the participants' basic assumptions about visibility improvements in general and the Grand Canyon in particular. Later focus groups paid particular attention to the Grand Canyon photographs to be used in the in-person surveys. The telephone surveys were conducted with random samples of Phoenix and Chicago residents. The questions in the telephone surveys were used to assess the representativeness of our focus group findings and to begin developing the structure of an in-person survey. While telephone surveys do not permit the use of photographs, they are quick to implement and proved to be very useful for the exploratory purpose for which they were used. The final activity during the preliminary assessment phase of our research was a small in-person survey we conducted in Provo, Utah. The instrument for this survey was an adaptation of the phone instrument modified for use with Grand Canyon visibility photographs. The results of this work suggested that a successful in-person survey based on days of visibility change in different seasons with a fairly extensive set of photographs was possible.

We also began to develop an instrument specifically for a larger in-person survey starting from the questionnaire developed during the assessment phase of our work. This instrument went through numerous drafts, a couple of small three-person focus groups, and a series of test interviews. We contracted with the National Opinion Research Center at the University of Chicago, one of the nation's leading survey research organizations, to conduct a pretest and pilot study.

The pretest took place at the end of June. After a day long interviewer training session for four NORC interviewers and their supervisors, the interviewers carried out a 22 respondent in-person pretest in four Chicago area neighborhoods under field conditions. A day spent debriefing the NORC interviewers and working with NORC staff resulted in substantial modifications that made administration of the survey more manageable and improved the clarity of the questions and the flow of the interview. The pilot instrument was administered to subsamples in St. Louis and San Diego during late July and early August. The average interview length was approximately 35 minutes.

Comparison of Questionnaires

The GCVB pilot instrument and the CR instrument both ask an extensive set of questions about park visits, awareness of visibility, and willingness to pay for changes from one level to a greater level of visibility. Both studies have respondents examine a series of photographs which depict different changes in visibility levels. The language of the two surveys for several of these questions is identical.

However, the GCVB pilot instrument differs from the CR instrument in a number of important respects. One obvious respect is the survey methodology; the pilot instrument is for an in-person survey whereas the CR Survey was a mail survey. In comparison with mail surveys, in-person administration offers the researcher a greater amount of flexibility in the kinds of questions that can be asked, the skip patterns* that can be managed, and the types of visual aids that can be employed to communicate complex information to the respondent. It also permits the researcher to control the pace and sequence of the interview whereas the mail survey respondent has control of these factors. Among the various differences in the wording of two survey instruments, three are particularly important.

First, the pilot instrument focuses strictly on willingness to pay for visibility improvements in the Grand Canyon whereas the CR Survey focuses on all Southwest parks. The CR Survey had respondents value visibility improvements in a number of Southwest parks *including* the Grand Canyon while the pilot instrument focuses entirely on the Grand Canyon. The only question specifically on the Grand Canyon in the CR Survey asked respondents what percentage of their willingness-to-pay to improve visibility in Southwest parks should be spent on the Grand Canyon. We believe the CR Survey's method for ascribing values to Grand Canyon visibility improvements is too casual for use in making an important policy decision since the respondents were not informed they would be required to give the percentage of their allocation to GCNP and they were not required to allocate their overall valuation among the several parks in the Southwest. If the Grand Canyon and the other Southwest parks are substitutes (Hoehn and Randall, 1989)[8] as we believe they are, then concentrating on the Grand Canyon exclusively, rather than on all the Southwest parks and apportioning some fraction of that willingness to pay to the Grand Canyon, other things being equal, should result in higher values for Grand Canyon visibility improvements.[†]

Second, the pilot instrument uses a conceptual framework and visual aids that approximate to a greater extent the type of Grand Canyon visibility improvement that may result from a BART action; the CR Survey limits itself to improvements described by relatively small photographs that depict large changes in average yearly visibility. This difference has two components. The first is the framework used to describe the changes. The CR Survey employed the concept of changes in average annual visibility. Respondents were shown a set of four photographs which they were told represented the current average yearly visibility level for the Southwest Parks and three other visibility levels, one worse and two better.

Our previous research on visibility values and our Grand Canyon focus groups and other survey development research lead us to reject the average annual visibility change approach. Although it has the apparent advantage of simplicity -- one photograph or number represents the visibility level of an entire year -- people do not find the average annual visibility level concept very meaningful. In addition, it tends to convey the erroneous impression that most days in the Grand Canyon have similar visibility conditions. The Grand Canyon is particularly unsuited to the notion of a change in average visibility because of its natural extremes in visibility and because of the strong influence of weather on visibility. Yet the average person may harbor preconceptions that the Grand Canyon does not ever suffer weather-caused visibility degradation; the pictures that most people see of the Grand Canyon rarely exhibit any visibility impediments. The visibility impairments attributed to NGS occur during times of high humidity,

*Skip patterns refer to directions in the questionnaire to control the flow of questioning by directing the respondent to the next question; which question is the next question depends on the response to the current question.

†At this point, we are not making any judgement as to whether a larger Southwest Parks policy should have been offered to respondents, only that by not doing so our estimates, if anything, should be higher than that of the RCS Study.

the same conditions during which weather-caused visibility degradation occurs. Thus the two cannot be separated in a questionnaire designed to elicit WTP values.

To arrive at the optimal balance between a simplicity which seriously distorts reality and a complexity which overwhelms the cognitive capacity of human respondents is very difficult. The pilot instrument attempts, without overloading the respondents, to convey a greater level of complexity by conveying three types of distinctions: (1) two seasonal periods, winter and summer, (2) three kinds of visibility conditions, and (3) individual days. Combining the first two distinctions yields six types of days -- summer high visibility, summer medium visibility, summer low visibility, winter high visibility, winter medium visibility, and winter low visibility. One of six photographs exhibits each of these types of days. While six types of visibility days necessarily simplifies reality a great deal, they nevertheless describe a much greater range of variation than does the CR Survey which describes change in terms of average annual visibility. Our respondents are also informed about the distribution of these different types of visibility days over the two seasons. Care was taken to ensure that the pictures shown for comparison were for the same times and same seasons, eliminating another problem in the CR Survey. We believe that, in spite of the additional complexity of our approach, respondents find it more plausible and therefore easier to understand.

The respondent is shown two photoboards, one for the distribution of three levels of visibility during the seven-month summer and one for the distribution of three visibility levels of the five-month winter. Each photoboard contains two sets of photos. The left set shows the current visibility levels, each photograph representing a certain number of days of the year. For each photograph of the left set, a photograph in the right set, the improved set, shows the improvement in visibility if most sources of pollution could be controlled. Those facts are presented to the respondents on the photoboards and in the text of the survey. Figure 1 displays a winter photoboard.

For the summer photoboard, the upper two photographs illustrate the current and improved visibility for 120 days; the center two photographs illustrate the current and improved visibility for 70 days; and the bottom two photographs illustrate the current and improved visibility for the remaining 20 days of the seven-month summer. The top pair of photographs illustrate the highest level of visibility and the bottom level the worst visibility. These relationships are illustrated in Table 1.

For the winter photoboard, the upper two photographs illustrate the current and improved visibility for 50 days; the center two photographs illustrate the current and improved visibility for 50 less clear days; and the bottom two photographs illustrate the current and improved photographs for the remaining 55 days of worst visibility. One of two versions of the pair of 55 worst day photographs were assigned to respondents to control for the possible effect of cloudiness and thereby avoid biasing the results in the event respondents reacted to the particular types of clouds shown in those photographs.* These relationships are illustrated in Table 2.

While both the CR and GCVB pilot instruments use photographs, they differ in focal length, size, and finish. The CR Survey used photographs 3 x 5 inches in size which were reproduced by a color printing process. They were taken by a camera with a 135 millimeter telephoto lens which focuses attention on distant details rather than presenting the view that a visitor might see. The GCVB photographs are 8 x 12 inches in size, printed on glossy photo paper. They were taken by a camera with

*The visibility ranges for the 55 winter weather days are weighted averages of the ranges of two sets of pictures each of which was shown to half the sample. Statistical tests showed no significant differences in willingness to pay responses to any of the five programs due to the version of the winter weather day shown to the respondent.

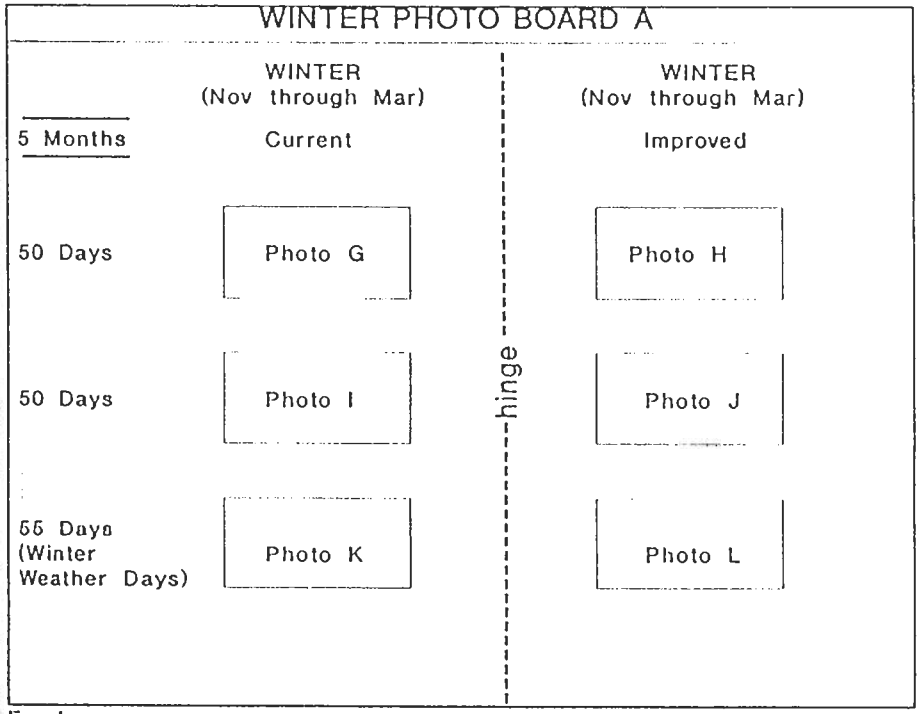

Figure 1

a focal length of 50 mm and depict the view a visitor would see. These characteristics of the GCVB photographs make them superior for communicating visibility levels. Our focus group participants and pretest interviewers strongly supported the use of photographs this size in preference to the same photographs in a 7 x 10 inch format because they thought the larger photographs did a much better job of showing Grand Canyon visibility.

Many of the respondents to the CR Survey may have had difficulty perceiving the visibility differences they were asked to evaluate. If so, the respondents would have been very sensitive to cues conveyed by the wording of the questionnaire. This may explain, for example, the otherwise puzzling fact that the CR Survey respondents were willing to pay a sizeable amount more for the visibility improvement from photograph C to photograph A than they were for the visibility improvement from photograph C to photograph B, even though the light extinction coefficient for picture B was slightly smaller than that of picture A.* In spite of being told that they were being shown summer pictures, they were effectively asked to value going from a hazy summer day to a clear winter day. Because winter

* The CR Survey showed respondents pictures A, B, C, and D, of which C represented current average annual visibility. They asked respondents to value changes from C to B, C to A, and to value preventing the change from C to D.

Table 1
COMPARISON OF CURRENT AND IMPROVED SUMMER VISIBILITY

Number of Days	Current Visibility (km)	Improved Visibility (km)	Average Daily Improvement
120	205	325	59%
70	111	243	119%
20	48	170	254%

Table 2
COMPARISON OF CURRENT AND IMPROVED WINTER VISIBILITY

Number of Days	Current Visibility (km)	Improved Visibility (km)	Average Daily Improvement
50	355	435	22%
50	98	205	109%
55	16	92	475%

colors are strongly preferred, this change in lighting conditions due to the season may have provided an upward bias to the resulting WTP values.

A third important difference is that the GCVB pilot instrument makes a much greater effort to get respondents to restrict their valuations to visibility changes. Indeed, the CR Survey took an opposite tack by suggesting to respondents in an earlier question that air pollution was responsible for damage to vegetation and historic structures and the acidification of lakes and streams. Furthermore, in the preamble to the valuation questions, the respondents are told that others would be asked to value the reduction of air pollution related effects on health, vegetation, and visibility.

Several additional differences between the GCVB pilot instrument and the CR Survey instrument deserve comment because they may have the effect of lowering the willingness to pay in the GCVB Study. In each case, design choices in constructing the pilot instrument were made on the basis of what seemed to be compelling reasons that this was a more correct design choice than the alternatives.

First, instead of using the CR Survey language which informed respondents that the visibility improvement would last forever, the pilot instrument informs respondents that the power plants responsible for the visibility impairment which is being corrected are scheduled to be in operation for 20 years. Thus, they are buying a good for 20 years after which time any visibility impairment due to the plants would end with the operation of the plant, and the visibility improvement would then be free. To the extent that respondents in the CR Survey thought they were buying a permanent visibility improvement and, more importantly, that these programs were the only opportunity to buy that improvement for future generations, our willingness to pay numbers should be smaller.

Second, we have changed the payment vehicle from higher prices and taxes to higher utility bills. If the policy being implemented is closely tied to a particular payment vehicle, as in the case of a NGS BART action, that payment vehicle should be used. The use of a higher prices and taxes vehicle is

justified when there are a multitude of ways in which respondents would actually pay for the good as in the case of national water quality. Such a payment vehicle would be appropriate in the Grand Canyon case only if a more comprehensive policy was being considered. Since any improvement in visibility resulting from an NGS BART action would be financed by higher electric bills, an electric bill payment vehicle should be used. It is possible that its use, rather than that of the higher prices and taxes payment vehicle, might lower the willingness to pay responses because of the immediate and direct impact of higher electric bills and because of resentment toward electric utilities. However, a rough comparison between the two payment vehicles in Phoenix and Chicago telephone surveys did not reveal any substantial differences between the two payment vehicles. Most respondents were unwilling to pay anything through either payment vehicle.

Third, unlike the CR Survey which begins with questions about national parks, our survey instrument begins with a set of questions that put the visibility improvements into a larger context by asking respondents about a set of issues, several of which are unrelated to the Grand Canyon. This type of opening question is frequently used by survey researchers to put respondents at ease and to encourage them to realize that the particular good that is the subject of the survey is but one of a much larger group of public goods. This approach mitigates to some extent the tendency of some respondents to assume that the good being valued must be valuable since so much money and trouble is being expended to get his or her views about it; this "importance bias" should be avoided in any study that attempts to obtain a credible benefits estimate using the CV methodology.

A fourth difference concerns the elicitation methods used in the two surveys. The CR Survey instrument uses the payment card method which, in their case, has the respondent choose one amount from a list of 28 amounts ranging from $0.00 to "More than $750." Ninety percent of the amounts on the CR Survey payment card are larger than the amount needed to justify the NGS BART decision. The configuration of this card is likely to have put pressure on the respondent who does not have any value for the improvement to circle a positive amount. In the pilot survey we use the open-ended method in which the respondent arrives at a dollar value with no prompting of any kind from the interviewer. While we have been strong proponents of the payment card method, its use requires that most respondents be willing to pay something. Its use also assumes that the willingness to pay amounts are distributed over a reasonable range rather than concentrated at a single value, such as zero.

The GCVB visibility valuation exercise consisted of several steps. The respondent was first asked to rank their preferences for the five different programs without considering costs. Then the respondent was asked which of those programs they would be willing to pay something extra for each year. This amounts to a dichotomous choice for willingness to pay of zero versus more than zero. If the respondent was unwilling to pay for any of the programs, he was given a chance to reconsider. Respondents not willing to pay anything for any program were asked to explain. For those respondents who would vote for any of the programs, the next step was to ask how much that the program could cost them before they would vote against it. Respondents who gave non-zero amounts were reminded that they were only obtaining visibility improvements and asked if that would influence the amount that they gave. Those respondents who said yes were asked to give revised amounts for just visibility. However, our focus group work and the Chicago pretest indicate that some respondents will not accept that the only effect of decreasing haze is to improve the visibility. Therefore, some willingness-to-pay responses may be biased upward. Respondents who were willing to pay non-zero amounts for programs were asked to explain what about those programs made them worth it; and respondents not willing to pay anything for programs were asked what about those programs made them not worth anything. This multiple step, yes/no, open-ended valuation, and explain format is appropriately sensitive to small values while avoiding the obvious threat of compliance bias (Mitchell and Carson, 1989: 237)[4].

PILOT STUDY RESULTS

A pilot study using the developed instrument was conducted at two sites, St. Louis and San Diego, from July 20 to August 2, 1990.[1] Since the intention of the pilot was to test the instrument under the type of field conditions that would prevail in a national survey, those sites were chosen to provide two diverse settings. We used an economical sampling frame based on random assignment of blocks and a quota scheme based on age and sex. Our quota-based sampling plan resulted in a random sample which looks reasonably representative of the population. We do not suggest, however, that our results should be extrapolated to the national population, merely that unlike the CR Survey, we measured the relevant visibility changes and that our results differ very substantially from those of the RCS Report.

In any telephone or in-person survey, interviewers necessarily play a key role in motivating respondents to cooperate and in eliciting information. This role must be carefully prepared and monitored to ensure the interviewers play their role in a neutral manner. At the conclusion of the interviewing, one of the researchers and the NORC staff conducted full day debriefing sessions in St. Louis and San Diego, attended by the interviewers and the local NORC supervisors. NORC headquarters personnel also attended the St. Louis debriefing. The debriefing probed the interviewers' experiences with the instrument, the problems encountered, and their suggestions for changes. Every portion of the instrument was systematically discussed.

The interviewers recorded all comments made by the respondents on the instrument itself. These verbatim responses are an important part of the survey results. They provide insight to what respondents were thinking and allow a much deeper interpretation of the quantitative findings. In some instances they suggest ways in which particular survey questions should be changed to avoid misinterpretation or confusion.

A large number of the respondents in our sample had visited the Grand Canyon: 38% in St. Louis and 54% in San Diego. Of the respondents who had been to the Grand Canyon, over half had only made one visit. As one might expect, San Diegans were more likely to have made multiple trips. Less than 20% in both samples had made more than two trips. If the respondent had visited the Grand Canyon, their last trip was on average a little over 10 years ago and the distribution of last trips was fairly uniform across years. Most trips appear to have been made in June, July, and August with only 3% of the last trips made by respondents in our St. Louis sample taking place in the five month November-March winter period and with only 10% of the last trips made by San Diego respondents taking place in that period. About 45% of the respondents from both samples rated visibility during their last visit to the Canyon as very clear. A little over 15% suggest that visibility was somewhat clear or not clear. Only 3% of the respondents had not visited the Grand Canyon or seen pictures of it. Approximately 30% of the St. Louis sample and 55% of the San Diego sample indicated that they were very likely or likely to visit the Grand Canyon in the future.

Virtually all of the respondents were able to chose the summer photograph with the best visibility; the correct visibility ordering on the picture boards had to be pointed out to only 4 respondents. Strong support was expressed for the environmental movement in both samples with the suburban St. Louis sample being somewhat more supportive (30% active; 61% sympathetic) compared to the San Diego sample (24% active and 46% sympathetic).[*] San Diego had a larger percentage of respondents who were

[*]Question 33 of our pilot study instrument used a four point scale: actively supports, sympathetic, neutral, and unsympathetic.

neutral or unsympathetic toward the environmental movement. Roughly 40% of both samples thought of themselves as outdoor people.

Valuation of Visibility Improvements

Estimation of willingness to pay for each of the programs requires a number of steps. For any particular program a respondent could give a zero, a positive amount, or a don't know. If they gave a zero, a determination must be made if it is a "protest zero". If a positive amount is given, one must check to see if a revision of that amount was made in Q28, where the respondent was reminded that the programs would only improve visibility. In a few cases, a respondent failed to give an amount for a program that was ranked higher, or should logically have been ranked higher, than a program they gave an amount for. In these cases, the value for this program was set to the amount given for the lower ranked program.

Don't knows tended to be of two types. One type consists of respondents who basically have no idea how much a visibility improvement program is worth to them and have no idea how to consider the issue; these responses are typically from the elderly and less well-educated. The other type of don't knows comes from the those respondents who say that they can not make the decision without more information, typically more information than could be provided in a contingent valuation survey. These responses tend to come from the better-educated and from specific occupations. This percentage of don't knows is low by comparison to most contingent valuation studies. The don't knows are excluded from our analysis of willingness to pay amounts here.

Classification of the zero responses into true zeros and protest responses is always a difficult and somewhat subjective exercise. If actually given the opportunity to vote on any of the visibility programs, many of these respondents would likely vote "no" so that removing all of these zero responses from the sample may artificially inflate the willingness to pay observable in an actual referendum context. We divided the zeros into four types. Those individuals who indicated that they did not have any real desire for the visibility improvement or that financially they could not afford to pay anything for the program were classified as true zeros. The other zeros were classified into three types: clear protests, possible protests, and likely true zeros. A clear protest was an individual who indicated that they wanted one or more of the visibility improvements but thought that it should be paid for in some other fashion. We classified as possible protests those respondents whose verbatim answers indicated that they did not like some aspect of the scenario but whose verbatim answers did not address whether they thought that one or more of the visibility improvement programs should be undertaken. Typically, these respondents indicated that the government wastes money, that the program should be paid from cuts in defense or the space program, or that the electric utility company could not be trusted. Likely true zeros look like possible protests in their complaints about the government and electric utilities. They differ in that they also included comments which suggest that they did not find the visibility improvement programs to be important relative to other problems faced by the government.

Of the 16 respondents who revalued their amounts after being reminded that they were only getting a visibility improvement, about one-third changed their amounts to zero, about a third reduced their amounts by half, and the other third changed their amounts in a somewhat erratic manner including one respondent who revalued upward and one who changed to a don't know. Eight respondents did not change their amounts but indicated that they believed that they had to be getting other types of benefits in addition to Grand Canyon visibility improvements. Thus only 12% of the sample appear to have not understood or believed the statements earlier in the questionnaire explaining that they would only be getting visibility improvements in the Grand Canyon; this percentage is a dramatic reduction from that

of the CR Survey where at least 58% of the sample allocated part of their willingness to pay for something other than visibility improvements.

The result is a set of yearly WTP estimates for the five programs. Table 3 displays the median, two frequently used robust estimators, the 5% trimmed mean and the 10% trimmed mean (Huber, 1981)[9], the mean, the standard deviation, the range, the percent of zero responses, the number of valid bids, the number of clear protest zeros, and the number of don't knows. Using the 5% trimmed mean as a crude upper bound and the 10% trimmed mean as a crude lower bound on the point estimate for the true, unobserved willingness to pay, the combined summer and winter program falls into the range $16-$20, the summer program $8-$11, the winter program $1-$3, the 20 winter day program $0.02-$0.50, and the 10 winter day program $0.00-$0.49. However, these estimates are significantly lower than those of the RCS Study, by a factor of 2 to 3 for the combined programs and by at least an order of magnitude for the 10 and 20 winter weather day programs. Formal confidence intervals show the willingness-to-pay amounts for the 10 and 20 winter day programs to be virtually indistinguishable from zero.

Table 3
VISIBILITY VALUES - PILOT STUDY

Program	Summer and Winter	Summer Only	Winter Only	20 Winter Weather Days	10 Winter Weather Days
Number of Valid Bids	183	182	182	181	181
Range	$0-360	$0-300	$0-150	$0-100	$0-100
Median	$10.00	$0.00	$0.00	$0.00	$0.00
10% Trimmed Mean*	$16.15	$8.11	$1.25	$0.02	$0.00
5% Trimmed Mean**	$20.20	$10.51	$2.92	$0.50	$0.49
Mean	$27.78	$15.71	$6.34	$2.38	$2.28
Standard Deviation	$50.04	$33.82	$18.89	$10.32	$10.08
Percent Zero	40%	58%	79%	89%	90%
Total Sample Size	202	202	202	202	202
Clear Protest Zeroes	10	10	10	10	10
Don't Knows	9	10	10	11	11

*18 outliers trimmed off each end of the distribution; **9 outliers trimmed off each end

A major issue in this analysis is the choice of correct summary statistic of willingness to pay for each program. The median is the statistic of choice if one adopts a majority voting rule. In this case, the combined winter and summer visibility program receives a median value of $10. All of the other visibility programs have a median value of $0, and therefore none of these programs would seem likely to receive majority approval for implementation.

The mean is the correct statistic under traditional welfare economic theory if one is willing to ignore distribution consequences, that is, if one is willing to accept a program of which all the benefits may be enjoyed by a few and the costs borne by the rest of society. However, the raw sample mean WTP should not be used as measure of benefits for any of the Grand Canyon visibility improvement programs. In contingent valuation studies which elicit continuous WTP payments, this statistic is known to be biased upward, sometimes by an order of magnitude or more. No major contingent valuation researcher has ever used or proposed the use of the raw sample mean WTP as a measure of the benefits of a program*. The reason is straightforward: like all data from surveys, a certain percentage of the data is "bad". By "bad", we mean that an observation's value for a variable is "missing", at variance with known facts, or inconsistent with behavioral models based on responses to other questions or economic theory. Government agencies such as the Census Bureau which collect vast amounts of data routinely use a variety of imputation techniques and consistency checks before the data is summarized and released as official statistics. The ordinary mean can be grossly distorted by a very small number of outliers such as the case here. A few outliers cause the ordinary mean to be anywhere from 1.5 to 8 times the size of the 5% trimmed mean, the larger distortions coming in the winter visibility improvement programs. Furthermore, because the gross errors on the small side are bounded by zero, the use of the sample raw mean WTP will almost always overstate the actual population WTP, often grossly so, by an order of magnitude or more.

We have taken several different approaches to estimating the mean WTP for the 10 and 20 winter day programs for the population sampled for the GCVB study. *All* suggest mean benefit values for the 10 and 20 winter weather day programs of $0.50 or less per household, and most suggest values substantially less than $0.50.† The results of this analysis is summarized below in Table 4.

One approach is using a α-trimmed mean for which the α percent smallest and largest observations are deleted and the mean is based on the remaining observations. The 10% trimmed mean, the mean based on the central 80% of the distribution, and the 5% trimmed mean, the mean based on the central 90% of the distribution are provided in Table 3. In many types of data sets, trimmed means are more reliable statistical estimators of the true expected value than is the ordinary mean (Stigler, 1977;[13] Huber, 1981)[9]. Contingent valuation surveys have a pattern of gross errors which make the α-trimmed mean, particularly attractive. This family of estimators can be thought of as implementing the notion due to Alan Randall that contingent valuation surveys provide a solid core of usable responses. Continuous, or nearly continuous, contingent valuation WTP data from an open-ended question such as we used in our pilot study, from a payment card such as the CR Study used, or from a bidding game tend to have a very distinct type of gross error pattern first noted by Brookshire, Ives and Schulze (1976)[10]: some zeroes which are not true zeroes and some very large stated WTP amounts which the respondent is not really willing to pay. These zeroes tend to reflect a rejection on the part of respondents of the scenario proposed and, in particular, the belief (possibly strategic) that there is some responsible party which should and will instead pay. Very large WTP amounts may in some instances reflect a desire to please the sponsor of the survey or the interviewer, a perception by the respondent that he is getting something more than the scenario is actually representing, or the opposite strategic view that a high stated

*For discussions in early studies see, for instance, Brookshire, Ives and Schulze (1976)[10] or Rowe, d'Arge, and Brookshire (1980)[11]. For more general discussions see Mitchell and Carson (1989)[4] or Smith and Desvousges (1986)[12]. Note that the CR Survey discusses possible invalid WTP responses in their data set and adopts an extensive set of *ad hoc* procedures to identify and remove these observations. Carson and Ruud (1991)[13] provided an extensive discussion of this issue.

†For extended discussions concerning different approaches to estimating mean willingness to pay from this data set see Carson and Ruud (1991)[13] and Carson *et al.* (1991)[14].

WTP will help the program to be implemented but that some or all of the cost will not actually be passed on.

It is possible to raise some legitimate objections to the use of an estimator which symmetrically trims off some predetermined percentage of the largest and smallest data values. These objections usually take three forms. The first is the simple objection to predetermined "symmetric" trimming. Those holding this objection usually contend that one of the two tails of the distribution is not likely to contain a high percentage of gross errors. To meet this objection, we have looked at the results of several univariate robust estimate techniques, the Huber, the Welsch, and Tukey's biweight. These techniques do not impose the symmetric trimming like the trimmed least squares estimators, nor do they automatically drop or downweight observations. They all result in lower estimates for the 10 and 20 winter day programs than the 5% trimmed mean and are essentially identical to the 10% trimmed mean estimates. These estimates are lower because the zero WTP amounts do not look like gross errors to these estimators while the α-trimmed means are automatically trimming off α percent of these zero WTP amounts at the same time they are trimming off α percent of the high WTP amounts.

The second objection is that very large WTP amounts may not still look like potential gross errors after one has taken account of covariates which may indicate that these amounts belong to respondents with high incomes, high concerns about national parks and visibility, and a likelihood of visiting the Grand Canyon in the future. This possibility is usually examined in the context of a regression on the available covariates. We look at three different regression approaches. The first is the standard econometric approach proposed by Belsley, Kuh, and Welsch (1980)[16] which has been used in previous contingent valuation studies such as Smith and Desvousges (1986)[12]. This approach effectively drops observations with large studentized residuals. When these observations are dropped, most of predictor variables have the expected signs, and many are quite significant even in a data set of our size. The second regression approach is a (Huber, 1981)[9] robust regression which conditions on the WTP amount for the combined winter and summer visibility program when predicting the WTP amounts for the 10 and 20 winter day programs. The strength of this approach is that as long as the same covariates predict the strength of preferences for the year round visibility program it is not necessary to have measured those covariates in our contingent valuation survey in order to determine whether a large WTP amount for the 10 or 20 winter day program is likely to be a gross error. The third regression approach we have examined is the Box-Cox proportional errors correction method suggested by Irwin *et al.*[17] This approach allows for the possibility that respondents may make errors in giving their WTP amounts which are proportionate to the magnitude of the WTP amounts rather than errors which are independent of the size of the WTP amounts they give. This approach results in lower estimates of WTP if the error distribution is estimated to be right skewed such as the log-normal and the model indicates that there is a large random component to the WTP amounts. All three of these regression approaches give estimates which fall between the 5% and 10% trimmed mean estimates.

The third objection is to taking a mostly statistical approach to identifying gross errors rather than looking instead at whether the WTP amounts respondents give are consistent with simple tests of consistency with economic theory. The simplest of these tests is to require that a respondent with a positive willingness to pay amount for one visibility program is willing to pay more for another visibility program which clearly dominates it in the sense of providing more of those visibility benefits. What we find here is that a substantial fraction of the raw sample mean WTP is due to a very small number of respondents whose gave the same positive WTP amount for the entire winter program (and typically also

for the combined winter and summer program) as they did for the 10 winter day program.* Dropping these respondents also results in WTP estimates for the 10 and 20 winter day programs between the 5% and 10% trimmed mean estimates.

Table 4

Estimator	20 Winter Days	10 Winter Day
Median	$0.00	$0.00
Mean	$2.38	$2.28
5% Trimmed Mean	$0.50	$0.49
10% Trimmed Mean	$0.02	$0.00
Huber	$0.01	$0.01
Welsch	$0.00	$0.00
Biweight	$0.00	$0.00
Belsley, Kuh, and Welsch Outlier Deletion	$0.15	$0.13
Conditional Regression (Huber)	$0.19	$0.15
Box-Cox Proportional Errors Correction[12]	$0.02	$0.01
Theoretical Consistency Check Only	$0.41	$0.37

*We find the fact that over 40% of the respondents to the CR Survey contingent valuation gave the same amount to all of the visibility improvements very strong additional grounds to recommend rejecting their estimates as having little bearing on Grand Canyon visibility benefits. In contrast, only 4% of the respondents to the GCVB study exhibit this type of inconsistent behavior.

CONCLUSION

We believe our Grand Canyon visibility research (Balson, 1990,[7] Balson *et al.*, 1990[9], Carson *et al.*, 1991[14]) leads to a simple conclusion. The CR Survey is *not* capable of providing any evidence on the benefits of this BART action. The estimates in the RCS Report are driven by unverifiable and highly questionable assumptions and not by any empirical data. A decision with sunk costs (Viscusi, 1988)[11] of the order of magnitude of this proposed BART action should not be made on such a basis.

This may seem to be an extreme position. However, summarizing the problems with the CR Survey as we will do here should cause anyone to ask how reasonable it is to base a major BART action on a single study which has the following characteristics: (1) it asked respondents to value an irrelevant average visibility change from hazy summer conditions to clear winter conditions while incorrectly telling respondents the pictures were of summer days, (2) it gave respondents a one time chance to buy a permanent improvement in Grand Canyon visibility, (3) it displayed small 3" x 5" photographs to respondents and did not reveal that a telephoto lens was used to focus attention on distant canyon details rather than presenting the view that a visitor might see, (4) less than 1/3 of the respondents indicated that their WTP responses were only for visibility, (5) over 40% of the respondents gave the same positive dollar amount to all of the visibility changes they were asked to value, (6) the allocation of willingness to pay for visibility improvements in all Southwest parks to improvements in just the Grand Canyon is based on a single question and not directly tied to the visibility change being valued, (7) that bequest motivations appear to play a large role in the values given even though the BART action concerns a single plant with a finite and reasonably short life span, and (8) that a mail survey (with either a bad sampling frame or bad response rate, depending on how one wants to define the problem) of four states, is used to represent the U.S. population. Add to this list of discomforting characteristics, the RCS Report's inability to theoretically justify going from willingness to pay for small annual average visibility change to willingness to pay for large visibility changes on a small number of winter days and the fact that the value for a small annual change in average visibility is determined almost entirely from the functional form assumed for the valuation rather than any actual data in that region.

The GCVB Study is a less substantial basis for making a major BART decision than we feel comfortable with. Its limitations stem from its pilot study nature and from continuing uncertainty over what visibility improvement would result from the proposed NGS BART action. It does, however, value the type of change which might occur from a BART action through its use of large well-defined photographs presented to respondents in in-person interviews conducted by one of the country's most respected survey organizations. On most of the criteria by which contingent valuation surveys are typically judged, the GCVB Study can be judged quite successful. If the GCVB Study results had been based on a large sample of the U.S. population using a full probability sampling design rather than a quota sample of 202 respondents at two sites, we would have no hesitancy in making a concrete statement on the magnitude of the benefits of the visibility improvements valued. Even so, at this time, the GCVB Study provides the best and, indeed, the only reliable estimate of the benefits to the public of a change in Grand Canyon winter visibility event days.

REFERENCES

1. W.E. Balson, R.T. Carson, M.B. Conaway *et al.*, (1990), Development and Design of a Contingent Valuation Survey for Measuring the Public's Value for Visibility Improvements at the Grand Canyon National Park, Report to the Salt River Project,(Los Altos, CA, Decision Focus, September.)

2. L.G. Chestnut and R.D. Rowe (1990), Preservation Values for Visibility Protection at the National Parks. Report prepared for the U.S. Environmental Protection Agency and the National Park Service. (Boulder, CO: RCG/Hagler, Bailly Inc.)

3. R.D. Rowe, L.G. Chestnut and M. Skumanich (1990), Controlling Wintertime Visibility Impacts at the Grand Canyon National Park: Social and Economic Benefit Analysis. Prepared for the U.S. Environmental Protection Agency. (Boulder, CO: RCG/Hagler, Bailly, Inc.).

4. R.C. Mitchell and R.T. Carson (1989), Using Surveys to Value Public Goods: The Contingent Valuation Method (Baltimore: Resources for the Future/Johns Hopkins University Press).

5. R.J. Kopp, P.R. Portney, and V.K. Smith (1990), "Natural Resource Damages: The Economics Have Shifted after Ohio v. United States Department of Interior, Environmental Law Reporter 20:10127-10131.

6. R.T. Carson, R. C. Mitchell, and P. A. Ruud, (1990) "Valuing Air Quality Improvements: Simulating a Hedonic Pricing Equation in the Context of a Contingent Valuation Scenario," in C. V. Mathai, ed., Visibility and Fine Particulates (Pittsburgh: Air and Waste Management Association).

7. W.E. Balson (1990), A Review and Critique of the Applicability of Visibility Studies to a Navajo Power Plant BART Decision, Report to the Salt River Project (Los Altos, CA: Decision Focus).

8. J.P. Hoehn and A. Randall, (1989) "Too Many Proposals Past the Benefit Cost Test", American Economics Review 79(3):544-551.

9. P.J. Huber (1981), Robust Statistics (New York: John Wiley).

10. D.S. Brookshire, B.C. Ives, and W.D. Schulze (1976), "The Valuation of Aesthetic Preferences", Journal of Environmental Economics and Management, 3: 325-346.

11. R.D. Rowe, R. d'Arge, and D.S. Brookshire (1980), "An Experiment on the Economic Value of Visibility," Journal of Environmental Economics and Management, 7:1-19.

12. V.K. Smith and W.H. Desvousges (1986), Measuring Water Quality Benefits (Boston: Kluwer/Nijhoff Publishing).

13. R.T. Carson and P.R. Ruud (1991), Grand Canyon Visibility Benefits Calculations. Memo To William E. Balson of Decision Focus, Inc.

14. R.T. Carson, M.B. Conaway, and W.M. Hanemann *et al.* (1991), Comments on the Benefit Analysis in the U.S. Environmental Protection Agency's Proposed Navajo Generating Station BART Action. Submitted to U.S.E.P.A.

15. S.M. Stigler (1977), "Do Robust Estimators Work with Real Data," Annals of Statistics, 5:1055-1077.

16. A. Belsley, E. Kuh, and R.E. Welsch (1980), Regression Diagnostics: Identifying Influential Data and Sources of Collinearity (New York: John Wiley).

17. J. Irwin, W. Schulze, G. McClelland *et al.* (1990), Valuing Visibility: A Field Test of the Contingent Valuation Method (Report to U.S.E.P.A).

18. W.K. Viscusi (1988), "Irreversible Environmental Investments with Uncertain Benefit Levels" Journal of Environmental Economics and Management, 15:147-157.