Measuring Nutrient Reduction Benefits for Policy Analysis Using Linked Non-Market Valuation and Environmental Assessment Models

Final Report on Stated Preference Surveys

Authors

Daniel J. Phaneuf
University of Wisconsin

Roger H. von Haefen
North Carolina State University

Carol Mansfield and George Van Houtven
RTI International

Version Date: February 2013

ACKNOWLEDGMENTS

This project was funded by Grant #X7-83381001-0 from the U.S. Environmental Protection Agency, Office of Water. Information contained in this report represents the opinion of the authors and does not necessarily represent the views and policies of the Agency.

The authors of this report would like to thank a number of people for their contributions to the work presented in this report. First, we acknowledge research team members Dr. Ken Reckhow and Dr. Melissa Kenney for their contributions to the survey design and analysis that we discuss here. Dr. Kevin Boyle and Dr. John Whitehead served as peer reviewers for the survey instrument and provided valuable advice on the design. Ross Loomis, from RTI International, helped with the analysis and development of the spreadsheet tool discussed in the report. We also thank Dr. F. Reed Johnson, also from RTI International, for assistance in the experimental design aspect of the survey. Finally, we would like to thank James Glover and Emily Hollingsworth, from the South Carolina Department of Health and Environmental Control, for their help in developing the survey instrument.

ABSTRACT

This document summarizes the economic modeling component of the project *Measuring Nutrient Reduction Benefits for Policy Analysis Using Linked Non-Market Valuation and Environmental Assessment Models*. The project's overall objective is to provide an integrated protocol that will assist state water quality managers in one aspect of their efforts to set numeric ambient nutrient pollution standards for surface water. The specific focus is on measuring the dollar-denominated benefits of nutrient reductions as they pertain to recreation and aesthetic services. To accomplish this task, a mechanism is needed that links measured nutrient pollution (e.g., ambient nitrogen, phosphorous) to a qualitative ranking of water quality, which can then be tied to an economic model of valuation. In this technical document we describe module 2 in our project, which centers on (a) designing and fielding a survey that uses stated preference methods to value water quality attributes, and (b) estimating economic models that take as inputs the predictions from our water quality model (constructed in module 1) and produce willingness to pay estimates of quality changes. We provide detailed descriptions of how water quality was described to survey respondents, how we designed choice experiment and contingent valuation method questions, and how our Internet-based sample was drawn. We also provide a full analysis of the data and present the main models we use for policy analysis. A case study is also included that draws together all aspects of the project.

CONTENTS

LIST OF TABLES

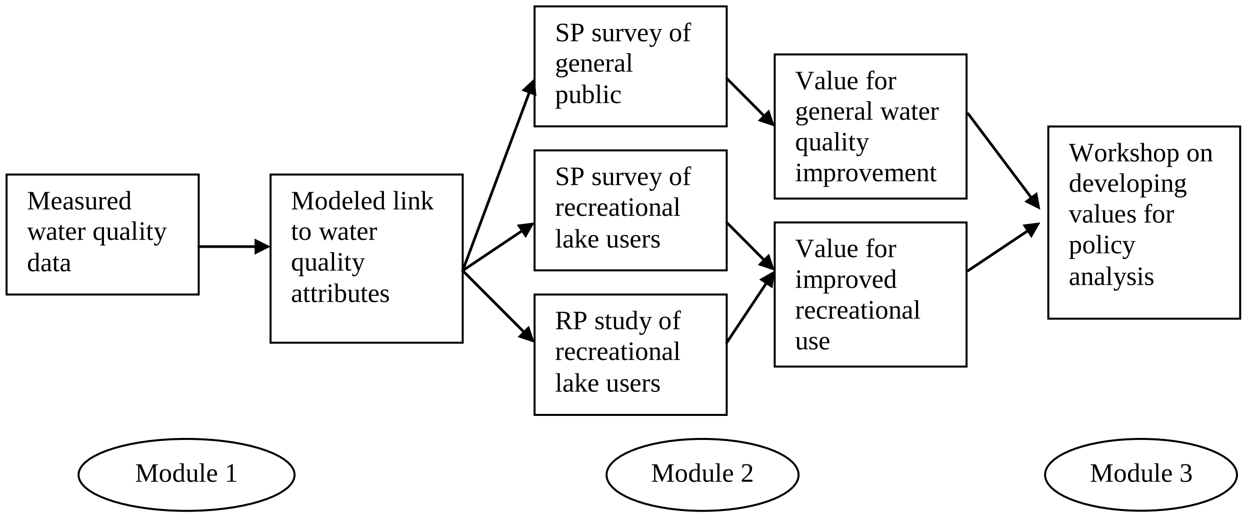
LIST OF FIGURES

1. Introduction

In this report, we summarize work on the project *Measuring Nutrient Reduction Benefits for Policy Analysis Using Linked Non-Market Valuation and Environmental Assessment Models*. The project's overall objective is to provide an integrated protocol for use by state water quality managers in setting numeric ambient nutrient pollution standards for surface water. The specific focus is on measuring the dollar-denominated benefits of nutrient reductions as they pertain to recreation and aesthetic services. To provide benefit estimates, a mechanism is needed that links measured nutrient pollution (e.g., ambient nitrogen, phosphorous) to a qualitative ranking of water quality, which can then be tied to an economic model of valuation. Figure 1 provides an overview of the project. Our objective in this technical report is to describe the research centered on the second part of the project (module 2): construction of an economic model that predicts the dollar-denominated benefits of improvements in water quality based on a stated-preference (SP) survey. This report complements our earlier report, *An Interim Report on Water Quality Modeling* (Phaneuf et al., 2009), which describes module 1 of the project.

To provide context for this effort, we first describe the motivation for the project and its overall structure. In 2007 the U.S. Environmental Protection Agency's Office of Water (EPA-OW) solicited proposals for research that would "aid States in their attempts to estimate monetary benefits associated with nutrient reductions as they strive to adopt numeric nutrient criteria into their State water quality standards" (EPA-OW, 2007, p. 2). The solicitation was motivated by the desire among state and federal managers to establish numeric (as opposed to narrative) water quality criteria and the realization that the costs of obtaining such criteria are more readily measurable than the benefits.

Figure 1. Overview of Proposed Project Modules



The request for proposals goes on to state:

> However, State agencies charged with developing standards and facilitating their adoption into state regulations often lack the staff time and funding required to do a complete analysis of benefits. To assist State lawmakers and the general public in being better informed, State environmental agencies need to be able to accurately characterize the economic value of environmental benefits associated with achieving water quality standards for nutrients. A thorough assessment of these benefits associated with numeric nutrient standards *would apply a production function approach, documenting the direct linkage between excess nitrogen and phosphorus in the water and a loss of ecological goods and services provided to society*, and provide a monetary estimate of benefits from restoring these services. (EPA-OW, 2007, p. 3, emphasis added)

In response to this solicitation, we submitted a project focused on the Southeast of the United States (EPA eco-region IX) that had three main objectives:

i. Development of a eutrophication production function whereby quantitative measures of ambient nutrient levels can be mapped to qualitative indicators of water body quality as reflected by its trophic status.

ii. Development of a revealed preference (RP) and stated preference (SP) framework for nonmarket valuation of the benefits of nutrient reductions that (a) links to the eutrophication production function; (b) is general in that the software, data sources, and analytical techniques are transferable to other regions and scalable for any policy question; and (c) is location specific in that the parameters of the benefit function can be calibrated based on local conditions and the local policy question of interest.

iii. Transfer of knowledge on how the framework can be applied for regulatory analysis via (a) a training workshop targeted at state-level water quality regulators and analysts and (b) distribution of software, data sources, and educational materials necessary for implementing the framework.

EPA-OW selected our proposal for funding, and work began in April 2008. In this technical report, we summarize our research on the second of these objectives.

Our starting point is the output from module 1 (see Phaneuf et al., 2009 for a summary of the results from module 1). As part of module 1 of the project, we examined models that would provide a mapping between a specific southeastern lake's measured water quality (e.g., total nitrogen and phosphorous, chlorophyll a, turbidity) and a qualitative/descriptive categorization of its nutrient pollution status. Module 1 built on a series of expert elicitations to link trophic states with numerical water quality data (Kenney, 2007). Our task in module 2 is to construct an economic model that takes a change in a lake's qualitative ranking as its input and provides a measure of value associated with the change as its output. For this we use the tools of nonmarket valuation, which we introduce in Section 2. In Section 3, we describe the development of the survey vehicle we used to obtain the data needed to estimate our economic model, and in section

4 we provide a summary of the information obtained. Section 5 presents the analysis for the two types of SP questions contained in the survey. Section 6 presents a case study designed to highlight the use of both the economic and water quality models for policy purposes. Section 7 presents our final discussion and conclusions.

2. Principles of Economic Valuation

We are interested in the economic concept of value for measuring the benefits to society of improving ambient water quality. Using the economic concept of value, benefits arise from how change in water quality improves the well-being of individuals (and households). Individuals' well-being depends on their preferences (i.e., the collection of their likes, dislikes, and viewpoints) and the income and time they have to satisfy their wants and desires. For example, if water quality in a lake improves, the well-being of people who like to visit that lake will likely increase. Similarly, if environmental conditions in local lakes improve, people who care about the state of the environment will likely experience an increased sense of satisfaction. There are many other examples of specific ways that individuals' preferences determine the extent to which they benefit or suffer from changes in lake water quality, and the size of the implied well-being changes can vary immensely among different people. Our task is to find an observable metric that reflects how improvements in water quality at specific lake sites change the well-being of people living in the states where the improvements occur. This metric is what we then use to express the benefits of improvements in monetary terms.

Since well-being is a subjective and abstract concept, we need a proxy for well-being that has quantitative meaning. A useful concept is maximum *willingness to pay* (WTP), which is the highest amount of money a person is prepared to part with to secure some outcome (see Freeman, 2003, for an overview of the welfare economic concepts linking WTP to well-being). For example, we might be interested in a person's maximum WTP to secure a change in water quality at a lake near his home. Suppose she has an income of Y and is willing to pay at most $\$X$ for the improvement. From this trade-off, we know that the level of satisfaction the person perceives must be similar between the baseline condition with income Y and the improved quality condition with income $Y - \text{WTP}$. Thus, WTP measures the value of the change, expressed in terms of what an individual would be prepared to give up to obtain it. Since improvements in water quality tend to be non-rival (many people can enjoy them simultaneously), the total WTP across the population of people affected by the change is a dollar-denominated reflection of the value the population holds for the change. Estimating individual and population WTP values using survey and statistical techniques is the applied objective of our study. In some instances we are also interested in distinguishing between *use value* and *nonuse* value. The former arises from peoples' direct interactions with the environment. For example, use value for improved water quality can arise from the improved recreation experiences that better quality enables. Nonuse value refers to the component of WTP not based on direct interaction. For example, a person can value water quality in lakes simply because it is important to him, even if he never intends to visit a lake for recreation. Likewise, a person can value an improvement because she wants to assure that healthy ecosystems are available for future generations.

Revealed preference (RP) and stated preference (SP) are two possible strategies for estimating WTP for changes in environmental goods such as lake water quality. RP uses observations of people's actual behavior to infer their WTP for a change in an environmental good, and therefore tends to be appropriate for measuring use value. The primary RP approach for valuing water quality is the travel cost model. The premise of this model is that trips to a recreation site (e.g., a lake for swimming) are costly because people need to spend time traveling and, usually, some money to reach the destination. If water quality is important to these people, they may drive farther (effectively pay more) to reach a site with better quality. By doing so, they reveal their WTP for water quality in the form of higher travel costs. By observing the actual choices among a sample of recreation trip takers, it is possible to estimate a WTP function for water quality as it relates to recreation behavior.

The SP approach presents survey respondents with detailed information about a specific environmental good (e.g., lake water quality in their home state) and then solicits responses to hypothetical changes in aspects of the environmental good. SP methods are quite flexible in that they are able to measure both use and nonuse values. We used two SP methods in this study: contingent valuation (CV) and choice modeling (also called conjoint analysis). CV directly questions people about their WTP. For example, after being introduced to the environmental good and a proposed (but hypothetical) change in the good, a survey respondent faces a question of the form *"Would you be willing to pay $X to have this change?"* If the survey is properly designed a yes answer shows the person has WTP $\geq X$, while a no answer suggests WTP $\leq X$. A full sample of responses allows estimation of a WTP function that depends on characteristics of the survey respondents. Questions of this type usually measure WTP in a way that includes both use and nonuse values, in that people can answer yes for a variety of motives. Choice modeling takes a less direct approach to estimating a WTP function. Respondents face choice situations in which they are asked to select their preferred option from two or more hypothetical possibilities. Each possibility consists of a bundle of characteristic levels. For example, the choice situation may be *"Suppose you are considering a visit to a lake. Which of the following two lake types would you choose to visit?":*

- Lake A: quality level = medium, travel distance from home = 20 minutes
- Lake B: quality level = high, travel distance from home = 40 minutes

When the survey is properly designed, the trade-offs people report making among the different attributes in the experimental design allow estimation of the value of changes in one of the attributes, relative to another. In this example, we can measure the value of water quality as reflected in the additional travel time people are willing to undertake to obtain a desired level of quality. The SP scenario in this case is related to the actual use of lakes, and so it will capture only the use value associated with improved water quality.

The effectiveness of an SP approach to valuing environmental goods depends critically on the details of how the survey is designed. In the next section we discuss the design of the survey we used to carry out the SP analysis for this project, which constituted the main approach we applied. We note first, however, that there is a large literature that has applied SP methods of various types to value water quality. Two examples of recently published work in this area include Herriges et al. (2010) and Viscusi et al. (2008). These papers use contingent valuation and choice experiment approaches, respectively, to measure values for freshwater quality improvements. Our study builds from and extends analyses of this type by linking the outputs from our eutrophication model to the commodity design aspect of our survey. We provide specifics about this approach in the following sections.

3. Survey Development

A key goal of our project was to design and execute a major survey of southeastern households to measure values held by residents for lake water quality. In this section we describe the survey design process. Throughout the design process we made use of standard survey research techniques. During the initial and intermediate stages of development, we held focus groups at three different locations over the course of 5 months. These locations included Raleigh, NC, in December 2008; Richmond, VA, in February 2009; and Charlotte, NC, in April 2009. To further assess the survey draft, we conducted one-on-one cognitive interviews with outside volunteers in September and October 2009. The survey was designed to be self-administered through a web interface. Between October and December 2009, we reviewed and tested the web version of the survey. We also commissioned peer reviews of the instrument from Dr. Kevin Boyle of Virginia Tech University and Dr. John Whitehead of Appalachian State University. In February 2010 we executed an online pretest of the survey, which involved 100 respondents. In the subsections that follow we describe in detail the interrelated steps we used to design the final survey.

3.1 Survey Objective and Overall Structure

In a broad sense, the survey objective was to gather data that would allow us to connect the water quality model estimated in module 1 of the project (which links numerical water quality measurements to an index of eutrophication) to economic values. Recall that our modeling effort in module 1 provided a function that linked chemical indicators of water quality at particular lakes to the qualitative/descriptive categories shown in Table 1 (and used for Kenney 2007). The chemical indicators such as levels of nitrogen and phosphorous are policy variables because numeric quality criteria can be set based on their values. The economic

Table 1. Trophic Status/Eutrophication Categories

Level	Water Clarity	Color	Algae	Nutrient Levels	Oxygen	Odor	Aquatic Life
1	Excellent	None	Very little	Very low	Very high	No	Very healthy, abundant
2	Good	Little	Little	Low	High	Little	Healthy, abundant
3	Fair	Some	Moderate	Moderate	Moderate	Little	Somewhat healthy, abundant
4	Poor	Noticeable	High	High	Low	Noticeable	Unhealthy, scarce
5	Poor	Considerable	Very high	Very high	Low to no	Strong offensive	Unhealthy, scarce or none present

benefits of better water quality, however, are not easily linked to these chemical values. Instead, people's preferences are more likely to be based on descriptive indicators of water quality expressed in lay terms. Given this, our primary concern was to examine preferences for water quality using the definitions and descriptions shown in Table 1. For the SP part of the survey, the challenge was to describe the water quality attributes in Table 1 in a way that was meaningful to the public but that conveyed the scientific understanding of the terms in module 1 of the project.

Economic theory and past research suggest that water quality affects people's utility through a variety of pathways. Water quality has been linked to recreation choices (Phaneuf, 2002), the value of lakeshore property (Leggett and Bockstael, 2000), concern for ecosystem health (Farber and Griner, 2000), and the cost of the public drinking water supply (Olmstead, 2010). For the public, the aesthetic characteristics of the water body and the quality of aquatic habitat are the characteristics most directly affected by nutrient levels. Many of the lakes in the Southeast, especially the large reservoirs, provide important recreation opportunities. Based on the goals of the project, we decided to focus on measuring the value of water quality changes for individuals who use lakes for recreation and the value of changes in water quality in the state as a whole for the general public.

The most natural format for eliciting the value of changes in water quality for recreation visitors is to observe how water quality affects trip decisions. RP analysis uses data on actual trips taken, and modeling is therefore based only on existing water quality conditions. Thus, for the SP choice experiment, we decided to have respondents choose between alternative trips to lakes with different levels of water quality. To match the travel cost paradigm, the cost of a trip was expressed as the time it takes to get to the lake. The choice experiment questions focus on day trips. We selected day trips as the object of choice for three primary reasons. First, the travel cost model is best suited for day trips in which driving time is the main cost (Phaneuf and Smith, 2005). Second, as discussed in more detail below, choice questions based on day trips are easier for respondents because fewer trip characteristics need to be described. Finally, a day-trip model allowed us to focus on lakes within a small radius around the respondent's home. Although the day-trip format has advantages, the result is that our estimates will not reflect the value of improvements in water quality for overnight trips.

We selected a contingent valuation (CV) SP approach to elicit the value the general public places on water quality improvements in the state. The CV question was presented in the form of a referendum in which the benefits and costs of lower nutrient levels were presented. Given the two SP approaches used, we decided at the beginning of the process that the survey vehicle would contain four main sections, presented to respondents in the following order:

a) <u>Recent recreation experience</u>. Gather information on whether the respondent has participated in lake-based recreation in the past year. If yes, solicit information about typical activities and the actual lakes visited. If no, solicit information on the future likelihood of a visit to a lake.

b) Water quality communication. Present information to respondents on how nutrient pollution affects lake water quality. Discuss how water color and clarity, fish populations, algae blooms, and the presence of odors can be used to classify lakes. Define and present five qualitative lake quality categories based on these dimensions that matched the descriptions in Table 1.

c) Choice experiment questions. Present lake recreation choice tasks to the subset of respondents who had visited a lake in the recent past or plan to visit in the near future. Design the choice tasks to solicit the trade-offs people are willing to make between distance from the lake and the water quality level when choosing a lake to visit.

d) CV question: Present the full sample of respondents (both those who answered the choice experiment and those who did not) with a CV question that solicits WTP for a program that improves water quality levels at lakes in the respondent's state.

As we discuss in detail below, the recreation and choice experiment sections are intended to assess water quality benefits as related to the recreational use of lakes, while the CV section is designed to measure the broader value of water quality improvements around the state. Section 4 discusses the fielding of the survey. Appendix A contains the final instrument on which the online version was based. No socioeconomic information was directly collected, because information of this type is available from Knowledge Networks (the marketing company we used to field the survey) for all members of their panel, independent of specific surveys. Each section of the survey was developed in an iterative fashion using focus groups, cogitative interviews, peer feedback, and discussion among the project team. We discuss the development process in the following subsections.

3.2 Recreation Section

In the recreation literature, trip activity typically correlates with an individual's WTP for water quality (see Herriges et al., 2010, and the works cited therein, for evidence of this from the Iowa Lakes Project). Reports on lake-based recreation provide information on the frequency and types of trips respondents take. Our objective in gathering recreation trip information was to obtain a relatively complete accounting of individuals' lake recreation trips during the last 12 months, while minimizing the burden to respondents of providing this information. This is a common objective in survey research designed to support travel cost modeling; thus, we were able to call on substantial past experience in designing this section (see Parsons, 2003, for more detail on data collection for recreation demand models).

The typical strategy is to present a "choice set" – the collection of sites for which trip-taking records are desired – and ask people to indicate which sites they visited and how many trips were made. Ideally, the choice set contains all the lakes a respondent might visit. Our design was made difficult by the large geographical scope of our survey: respondents were drawn from 8 states, including North Carolina, South Carolina, Virginia, Georgia, Kentucky, Tennessee, Alabama, and Mississippi. Thus, the number of lake sites we might consider presenting to people was comparatively large. Given this, our specific tasks were to (1) design skip patterns that would allow people to move quickly through irrelevant sites and (2) design choice sets that contain the major lake sites in all the states, while keeping lake lists small enough to view on a single screen page.

We ultimately settled on a telescoping skip pattern, which proceeded as follows:

- People were asked if they had taken any day trips to lakes for recreation in the last 12 months. Our previous experience suggests that 12 months is long enough to distinguish between frequent lake visitors, occasional visitors, and nonvisitors but short enough that recall problems should not be a significant issue.

- Trip takers were presented with a list of states within our broad study area and asked to indicate which states they had visited for their lake trips.
- For each state visited, respondents were presented with a map dividing the state into specific regions (e.g., western North Carolina, central North Carolina, and eastern North Carolina) and were asked to indicate the region(s) in which the lake(s) they visited was located. This was done for each state sequentially.
- For each region/state combination in which a trip occurred, the respondent was presented with a list of lakes and asked to mark those they visited. An option to write in a lake name was provided for cases in which the lake was not listed. Once the lakes for the region/state were obtained, a new prompt asked for the count of trips to those lakes. This was done for each region/state combination sequentially.

This strategy for obtaining trip counts was tested mainly via cognitive interviews and the pretest of the survey. Interview respondents reported having little trouble reporting their activities in this manner, and the pretest data set had few item nonresponse instances for this set of questions.

We relied on three sources to design the lake choice sets for each state/region that would be presented to respondents. As part of previous research, co-PI Dr. Kenney assembled an inventory of lakes in North Carolina that formed the basis for that state’s lake list. Her research also provided input into identifying the main recreation lakes in Virginia and South Carolina. For other states in the study area, our initial source of information was web sites maintained by state government agencies for tourism promotion and/or water management. Project team members with substantial experience with water issues in the Southeast reviewed these lists for completeness. These reviews produced ad hoc additions and clarifications on naming conventions. Finally, we used data from the 2008 National Survey of Recreation and the Environment (NSRE) to cross-check our lake lists. The NSRE is a large, country-wide survey of recreation behavior that includes a module summarizing freshwater recreation activity, including the names of particular destinations. We listed lakes in each state in our study area that respondents to the NSRE named as having visited and ranked these by the count of visits reported. We matched these data against our initial lake lists to be certain that we had not left out lakes that may be important recreation destinations. This effort led to the addition of a small number of lakes to our lists. Appendix B contains the final region/state-specific tables used to program the survey. Ultimately, our choice set included 1,117 named lakes across the states in our study region.

3.3 Water Quality Communication

As noted above, our survey plan involved using a choice experiment and CV to measure the trade-offs in time and money respondents would make for improved lake water quality. A critical component of any SP exercise is to define the environmental good in question and communicate how different quality levels of the environment might affect survey respondents. In our case, the environmental good is lake water quality, which is affected by nitrogen and phosphorous loadings. Recall that our water quality model links chemical and physical measures of water quality to the five-level eutrophication index presented in Table 1. The eutrophication index is based on seven traits of lakes that have ordinal, qualitative levels. For each value in the eutrophication index, the seven traits are set to levels consistent with what one would generally expect to find in lakes at that index value. The SP section uses five of the seven traits to elicit the public's preferences for improving water quality from one level to another by reducing eutrophication. We dropped the traits "nutrient levels" and "oxygen." The general public does not perceive nutrient and oxygen levels per se, but rather the effects of these levels on observable traits.

The water quality communication section of the survey begins with a short discussion of how excess nitrogen and phosphorous can affect water quality. The next few pages of the survey present the traits used to describe the level of eutrophication, with the objective of providing a survey-appropriate version of Table 1. Starting with Table 1, we needed a scheme for ranking the quality levels of different lakes that was accessible to survey respondents, based on perceptible features of the water bodies; scientifically accurate, appropriate for both the choice experiment and the CV question; and could be linked to the five levels shown in Table 1.

Table 2 displays our final definitions for the five categories of lake water quality, along with five perceptible traits that were used to describe each level. These traits are water color, water clarity, fish populations, algae presence, and odor presence. Note that each trait can take on three to five descriptive levels, so all trait levels do not vary over all categories. The categories A through E in Table 2 correspond to the levels 1 through 5 in Table 1. Note that the descriptions in Table 2 are abbreviated, and the survey instrument provides more explanation on each category and level (see Appendix A).

The process of moving from the conceptual objective to the final categorization shown in Table 2 involved several steps. The first was to establish whether water quality is likely to matter to people generally. This is important because one of the criticisms of SP survey techniques is that they create the preferences they are intended to measure. Focus groups are an important tool for examining the relevance of an issue as well as effective communication

Table 2. Abbreviated Water Quality Descriptions for the SP Survey[a]

CATEGORY	A	B	C	D	E
Color	Blue	Blue/brown	Brown/green	Brown/green	Green
Clarity	Can see 5 feet deep or more	Can see 2–5 feet deep	Can see 1–2 feet deep	Can see at most 1 foot deep	Can see at most 1 foot deep
Fish	Abundant game fish and a few rough fish	Many game fish and a few rough fish	Many rough fish and a few game fish	A few rough fish but no game fish	A few rough fish but no game fish
Algae blooms	Never occur	Small areas near shore; some years, 1–2 days	Small areas near shore; most years, 1 week	Large areas near shore; once a year, 2–3 weeks	Large, thick areas near shore; every year, most of summer
Odor	No unpleasant odors	1–2 days a year, faint odor	1–2 days a year, faint odor	3–4 days a year, noticeable odor	Several days a year, noticeable odor

[a] The survey instrument provided the respondents with more details on each category and level. Categories A to E correspond to levels 1 to 5 in Table 1.

approaches. The participant worksheets used for our three focus groups are included as Appendix C. Our first and second focus groups included tasks designed to establish whether water quality matters to people when considering lake recreation destinations. By prompting participants to list important features of lakes *before* mentioning the purpose of the focus group, we were able to examine the extent to which water quality indicators are salient in people's recreation choices. Along with obvious features such as distance from home, facilities, boat ramps, and swimming areas, a sampling of features people listed included the following:

- clean and neat generally—no litter near facilities, near-shore areas are debris free;
- pleasant scenery with healthy vegetation;
- quiet, a lack of crowds;
- water appears clean; no health worries about going in the water; and
- lack of pollution presence generally.

The focus group evidence suggests that water quality is a feature that is likely to matter to people in their recreation decisions. This provided assurance that water quality can be reasonably expected to enter preferences through recreation and that our choice experiment concept was valid.

The next step was to examine the ways that people judge the quality of lakes they might visit. Our hypothesis was that perceptible and intuitive features would be most relevant, rather than scientific measurements. In the first and second focus groups, we asked people to list features of lakes that they associate with cleaner and dirtier lake water. A sampling of clean-indicating responses included

- more fish and wildlife;
- clearer water; healthy looking color; and

- fresh and natural smells.

A sampling of dirty-indicating responses included

- stagnant water and rotting smells;
- surface scum, dead fish; and
- low water levels.

We used the responses from the focus groups, coupled with the descriptions from Table 1, to create descriptions of the five traits (color, clarity, fish populations, algae presence, odor presence) that we used to distinguish lakes of different quality. With the traits identified in the focus groups, input from Dr. Reckhow and Dr. Kenney, and the advice of outside water quality scientists were critical to our efforts to match our lay descriptions in Table 2 to the scientific understanding of the experts interviewed to create the water quality function in module 1.

In the survey instrument, each of the five traits is discussed separately. The survey instrument provided a short description of the trait, how it is affected by nitrogen and phosphorous, and the levels it could take. After each description, the respondent answered a question about the trait. The questions were intended to prompt the respondent to think about the trait and to break up the text in the survey, which helps keep respondents focused.

We first considered color. From Table 1, color ranged from none to considerable. The water quality scientists understood that the labels referred to the color associated with excess nutrients (primarily shades of green and brown). For the public, the actual color of water with varying levels of eutrophication needed to be described. In situations such as this, pictures and graphics provide an important complement to text in surveys. The NOAA panel[1] (NOAA, 1993)

[1] The National Oceanic and Atmospheric Administration (NOAA) panel was a panel of experts convened by the NOAA to review the use of CV surveys in natural resource damage assessments and to provide recommendations for best practice.

recommends including visual displays in addition to text, and in past studies we have always included pictures or graphs. Mathews et al. (2007) also emphasize the usefulness of visual aids in presenting information, both picture and graphs. Their discussion touches on the subjective judgment that must ultimately be employed in a decision about visual aids, and they stress the importance of pretesting to make sure that respondents do not misinterpret the visual aids.

We ultimately decided on four color levels: blue, blue/brown, brown/green, and green. In focus groups we started with five levels and provided participants with five photos of a single lake scene. The colors were digitally manipulated to show five color gradations with varying degrees of blue, brown, and green but were presented out of order. People were asked to identify the photo that best fit lakes in their area, and the subsequent discussion focused on rating which colors corresponded to the cleanest water. There was general consensus that blue was better and that gradations toward brown/green combinations indicated worse water quality. We settled on four levels of color (categories C and D both have brown/green color) because focus group respondents presented with five shades of color did not distinguish well between the middle brown/green categories. The survey instrument described the impact of nutrients on water color and presented photographs of the four levels (see Appendix A). The descriptions were somewhat complicated by the natural brown color of many Southeastern lakes. However, based on discussions with Dr. Kenney about how the water quality scientists she interviewed would have thought about color and the focus group results, the top category was designated as blue.

Like color, clarity is a visual trait of the water. In Table 1, clarity ranges from excellent to poor, with categories 4 and 5 both associated with poor clarity. For water clarity we decided on four levels: can see 5 or more feet, can see 2 to 5 feet, can see 1 to 2 feet, can see at most 1 foot. As in Table 1, the lowest two categories, D and E (levels 4 and 5 in Table 1), were both associated with the lowest level of clarity. The focus group discussion indicated that people were comfortable with this trait and understood what clarity meant. Again, the adjectives in Table 1 were translated into concrete descriptions with the help of water quality scientists to ensure that the levels used in the SP survey correctly matched with the scientists' understanding of the associated terms in Table 1. The main challenge was to suggest a tool people could use to visualize the differences mentally. Based on focus group discussions, the image of an angler's bait disappearing beneath the surface after some depth seemed the most effective.

Water color and clarity are visual traits that did not need to reference spatial or temporal dimensions to be effective, although, of course, they will vary depending on weather and other factors. The remaining three traits—aquatic life, algae and odor—were comparatively more complicated. In Table 1, aquatic life is described as ranging from "very healthy, abundant" to "unhealthy, scarce or none present." The focus group discussions suggested people were comfortable with the type and population of fish as an indicator of aquatic life. In the focus groups, anglers were familiar with the species of fish that thrive in lakes with better and worse water quality. Although nonanglers might have been less familiar with the names of the fish, they did understand that cleaner water supported greater species diversity, more highly valued fish species, and larger populations of fish. After testing descriptions of fish type and abundance in the focus groups, the participants could easily understand the distinction between game fish and rough fish. They also accepted that clean water was capable of supporting game fish populations (e.g., bass and crappies), while dirty water tended to be dominated by rough fish (e.g., carp). We, therefore, decided to describe fish habitat based on which types of species were most prevalent. The five levels we settled on include abundant game fish and few rough fish, many game fish and few rough fish, many rough fish and few game fish, and a few rough fish but no game fish. Based on focus groups and discussions with scientists, we used carp and bullhead catfish as examples of rough fish and bass, crappie, bluegill, and channel catfish as examples of game fish.

Describing the algae and odor trait levels posed special challenges because of the spatial and temporal variation in their outcomes. The algae levels in Table 1 range from “very little” to “very high.” Again, scientists can interpret these terms based on their knowledge and the information provided to them in the expert elicitation. For the general public, more detailed descriptions were needed. In the focus groups, people were shown a photo of an algae bloom and asked to discuss whether they had seen such a phenomena. Participants generally knew that a bloom signaled something was not correct, but few had observed a major bloom covering large sections of a water body. Participants also appreciated that blooms were seasonal and temporary. However, because most of the focus group participants had never seen a large algae bloom in lakes where they visited, including the algae attribute caused some people to think that water quality was better than it actually was because of the lack of surface algae blooms. We addressed this problem in two ways. First, the final descriptions for the algae attribute described both spatial and temporal variation in algae. These dimensions were combined into five levels: blooms never occur, small areas of algae occur in some years and last a few days, small areas of algae occur in most years and last a few days, small areas of algae occur every year and last several days, large areas of algae occur every year and last several weeks. Even with spatial and temporal variation, some respondents had difficulty incorporating algae into their decisions, because they assumed they could take care of this problem by not visiting the lake when there were algae blooms. Our concern was that respondents would be distracted by the algae attribute. As discussed below, the odor attribute had similar problems. To test for the important of these two attributes, we designed two versions of the water quality descriptions—one that included algae and odor and one that did not. This allowed us to examine the extent to which responses differed across the two treatments.

As with algae, odor proved to be a difficult trait to describe accurately. In Table 1, the odor levels range from “no” to “strong, offensive.” Although people appreciated that lakes can smell unpleasant, gaining consensus on how frequently this might occur in good versus bad quality lakes was more elusive. The focus group participants had never been at a lake that smelled anything close to “strong and offensive.” Similar to the algae attribute, individuals tended to think water quality at lakes must be better than it actually is, because they had never experienced any problem with odor. We decided on five levels for this trait: strong unpleasant odor several times a year, noticeable unpleasant odor several times a year, noticeable unpleasant odor 2 to 3 times per year, faint unpleasant odor 1 to 2 times per year, no unpleasant odors.

In SP surveys, designing the attributes and levels for the attributes requires balancing the cognitive burden of the survey instrument against the information needed to make a decision. This balance was our primary concern in communicating the attributes described above. During the expert elicitation research, Dr. Kenney conducted long, detailed interviews with water quality experts who understood the subtleties of water quality science, including the nonlinear and correlated structure of eutrophication indicators. Although our descriptions for the lay public cannot convey the same subtleties, our focus group and pretest work and our discussions with water quality scientists suggest respondents were able to understand how attribute levels related to the water quality index in ways that scientific specialists would broadly agree with. Thus, we are confident that the descriptions shown in Table 2 provide a solid basis for our SP questions.

3.4 Question Design

The survey questions needed for our SP objectives fall into three categories: survey questions designed to explain and reinforce the water quality categorization used, the choice experiment questions, and the CV questions. A set of questions were designed to familiarize respondents with the water quality traits and their levels individually (see questions 10.1 to 10.5 in Appendix A). For each of the traits—color, clarity, fish population, algae, and odor—the survey provided explanations on how nutrient pollution can affect the levels of the trait in lakes across their region. For each trait, individual respondents were asked to review the levels and then indicate which level they thought most closely corresponded to the lakes in their area. Ultimately, peoples' answers to these questions were less important to our analysis, but by asking people to consider lakes in their area we encouraged them to read through and think about each trait and its possible levels. For water color and algae presence, photos were also included to provide visual cues (see questions 10.1 and 10.4). After reviewing the water quality traits individually, the next set of questions (questions 11, 12, and 13 in Appendix A) asked respondents to think about lakes in their home state using all five of the trait levels simultaneously. Respondents were presented with information similar to Table 2 and asked to select the lake category that they believed corresponds most closely to lakes in their home state. In this way people gained experience in thinking of lake quality as consisting of a related collection of perceptible indicators. To reinforce this, the next two questions asked respondents to consider the remaining categories and indicate which among them had the next best level of correspondence with lakes in their states. With the water quality communication section so completed, respondents who were lake recreators or potential lake recreators completed the choice experiment questions.

Choice Experiment Questions

The intent of our choice experiment was to examine trade-offs people make between water quality and other scarce resources, within the context of recreation behavior. To accomplish this, we presented people with sets of lakes that were differentiated by their water quality levels and other relevant attributes (such as travel time from home) and ask which they would prefer on a given trip-taking occasion. We had to make several decisions before arriving at the final form of the choice questions, an example of which is shown in Figure 2.

Figure 2. Example Choice Experiment Question

		LAKE 1	**LAKE 2**
WATER QUALITY	WATER QUALITY CATEGORY	**C**	**B**
	COLOR	Brown/green	Blue/brown
	CLARITY	Can see 1–2 feet deep	Can see 2–5 feet deep
	FISH	Many rough fish and a few game fish	Many game fish and a few rough fish
	ALGAE	Small areas near shore; most years, 1 week	Small areas near shore; some years, 1–2 days
	ODOR	Faint odor, 1–2 days a year	Faint odor, 1–2 days a year
ONE-WAY DISTANCE FROM YOUR HOME		[30-minute drive]	[90-minute drive]
Which lake would you choose? *(check one box)*		☐ **LAKE 1**	☐ **LAKE 2**

These included the plausibility of the trade-offs we wanted to model, the subset of people who would answer the questions, the specific form of the choice scenario, the role of different activity options, the collection of attributes that would characterize the lakes, and the role of an opt-out or "no trip" option.

As Figure 2 shows, our choice experiment focuses on trade-offs between water quality and travel time in the selection of lake destinations. We used the first focus group to scope out the extent to which this is a trade-off people accept as realistic, and consider in their actual decisions. Our discussion indicated that people understand that better water quality can result in a better visit experience, but that the activity that is planned would condition the extent to which water quality matters. Our summary of the open discussion in the first two focus groups suggests participants would be willing to drive 30 to 60 more minutes to reach a destination with appreciably better quality, particularly if it were for swimming or angling or if a child would be in contact with the water. During our second focus group, we also examined how people would respond to explicit trade-offs between travel time and water quality as defined for the survey, presented within a format as in Figure 2. The discussion indicated that people had little trouble accepting the notion that such trade-offs were possible, and participants again reported a willingness to travel further to obtain better quality. From our focus group work we are confident that the trade-offs we are modeling are grounded in reality and that people accept the notion that such trade-offs may be relevant. We also learned that the extent to which people think about the distance/water quality trade-off may depend in part on the activity that is planned and the composition of the group.

Because our survey used a general population sample, we needed to take steps to ensure that only people for whom lake recreation was relevant would answer the choice experiment questions. We used answers to the recreation experience question (question 1) to screen individuals. If respondents indicated they had not visited a lake for recreation in the last 12 months, they were asked if they thought it was likely that they would visit one in the coming 12 months (see question 2). If respondents visited a lake or indicated that they might in the future, they were tracked to the choice experiment section. At that point they were presented with a scenario that begins:

> *Imagine the following situation. Sometime next summer, the weather forecast for the weekend looks good so you begin thinking about a day trip to enjoy your favorite lake recreation activity.*

Because the focus group evidence suggested activity and group composition may matter in how people make trade-offs, respondents were then asked to indicate what their main activity would be, whether they would travel alone or in a group, and whether the group would include children. Survey questions 14 and 15 (see Appendix A) show the specific wording we used and the options that people could choose from. Answers to these two questions allow us to condition our analysis on what people imagine to be the intent of their recreation trips. Based on the focus groups, we knew that respondents would probably imagine the activity they would do on their lake trip. We decided to have people state their activity and group composition, because it did not seem realistic to assign these features to their trip. In addition, explicitly asking activity and group composition would allow us to better control the variability in answers.

Presenting different lakes for survey respondents to choose from required decisions on what attributes of the lakes to present and vary. Choice experiments can generally be divided into two types: branded and generic. A branded approach in our context would be to name *specific* lakes near a respondent's home (e.g., Falls Lake and Jordan Lake for residents of central

North Carolina) and present real or designed quality levels at the lakes for people to select among. This approach has clear disadvantages for our purposes, in that people's choices will reflect their unobserved (to us) attitudes toward the lakes, thereby invalidating our measurement strategy. A generic approach, by contrast, presents people with unnamed options that are distinguished only by the attributes included in the experiment. Though this is a more abstract concept for the individual, it allows us to measure the trade-offs we are interested in while hopefully avoiding confounding individuals' attitudes about or experience with specific lakes. To characterize the generic lakes we elected to use only two attributes: the water quality category (consisting of the bundle of traits and their levels) and the travel distance to the lake. We could have included other attributes such as the presence of facilities, a swimming area, and parking lots, but this would have complicated the design and increased the amount of information people would need to process. Because these attributes are largely orthogonal to our objectives, we left them as implicit rather than explicit variables. Each choice question was introduced as follows:

> *Imagine that your two options are Lake 1 and Lake 2. The only differences between these two lakes are shown in the table below. Otherwise, they are exactly the same in every other way.*

Thus, we invited people to form their own images of what the lakes would be like outside of the designed attributes but asked that they not differentiate their images between the two options. Our probing during focus groups indicated that people were comfortable with this level of abstraction.

The final decision regarding the design of the choice experiment questions had to do with the use of an opt-out option, in which the person can decide not to choose one of the options. In general, the decision to include an opt-out option depends on whether the respondents could realistically opt-out of the choice in the real world and the goals of the study (see Hensher, Rose and Greene, 2006, or Holmes and Adamowicz, 2003). For example, in this study, respondents could decide they did not want to take a trip given the choices available to them. If respondents can realistically decide not to choose any of the alternatives, including an opt-out alternative provides respondents with a more realistic choice situation. In addition, if the goal of the study is to predict demand, omitting the opt-out option will likely bias the predictions. Holmes and Adamowicz (2003) favor the inclusion of an opt-out option. However, an opt-out option reduces the amount of information gathered, because the person does not provide a ranking of the designed options if they select no trip. To balance these two features, we used a two-step approach in which respondents first indicated which was preferred between the two generic lakes and then was given the option of visiting neither if the attribute levels were not to their liking.

Contingent Valuation Questions

Though the primary intent of the survey is to assess the value of lake water quality as it relates to recreation, we decided to add a more general valuation vehicle for two reasons. First, we expected that at most half of the respondents in a general population survey would be lake recreators or potential lake recreators. Given this we wanted an auxiliary vehicle that was general enough for all respondents to answer. Second, there may be instances when the value of a broader policy intervention—for example, at the level of the entire state—may be useful for state water quality managers. For these two reasons we decided to include a CV exercise that would examine respondents' general (as opposed to recreation-specific) WTP for broad-scope policies that would improve lake water quality across their entire state. Designing the CV section of the survey required a definition of the commodity to be valued, a strategy for explaining baseline and improved levels of the commodity, a specific policy project that would provide the improvement, a payment vehicle, and an elicitation method.

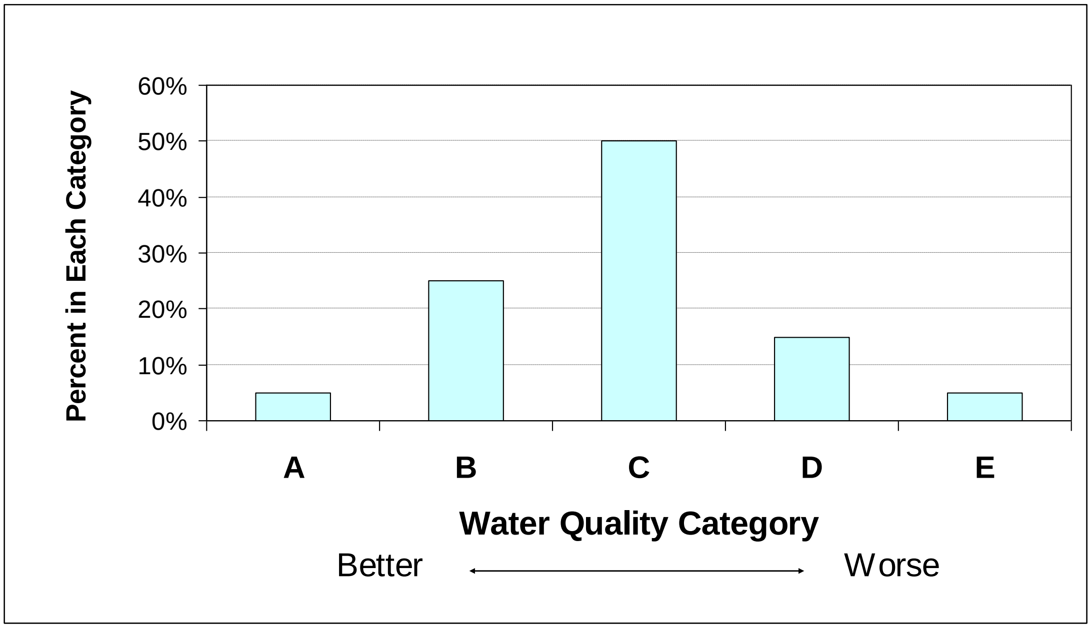
Our goal in defining the commodity for the CV exercise was to describe water quality using the five categories discussed above, but in a way that would be broader than the single-lake/single-trip focus used in the choice experiment. We settled on an approach in which the overall commodity would be the distribution of lakes across the five categories in the respondent's home state. Thus, our valuation task was to measure people's WTP to shift the distribution of lake water quality from baseline conditions to some hypothetical improvement. This required that we communicate a baseline condition and how a policy might change things. We used both textual and graphical descriptions to accomplish this. As is recommended (Mathews et al., 2007), we pretested the graphic carefully to make sure that respondents understood the graph. In particular, the CV section began with a written description as follows:

Information about water quality at public lakes is often collected and reported by state agencies. This information can be used to show the percentage of lakes in HOME STATE that are in each of the five water quality categories.

- *30% (3 out of every 10 lakes) are in one of the best two categories (**A** or **B**)*
- *50% (5 out of every 10 1akes) are in the middle category (**C**)*
- *20% (2 out of every 10 lakes) are in one of the lower quality categories (**D** or **E**).*

A graphic as shown in Figure 3 was also provided to give a visual image of the baseline conditions.

Figure 3. Graphical Description of Baseline Water Quality



After presenting the baseline, we described a generic program at the state level that would provide a general improvement in water quality across the state. Respondents were presented with the following program description:

> *Imagine that state agencies in charge of water resources in HOME STATE are considering a program to improve lake water quality. Because nitrogen and phosphorus come from many different man-made sources, there are many ways to control them. Under the program being considered, efforts to reduce nitrogen and phosphorus would be spread among many different groups. For example,*
>
> - *sewage treatment plants would have to install better treatment systems;*
> - *residents using septic tanks would have to inspect these systems for leakage;*
> - *towns and housing developments would have to install improved systems for managing water runoff from storms;*

- *farms would have to reduce fertilizer runoff from fields and improve the containment of animal waste.*

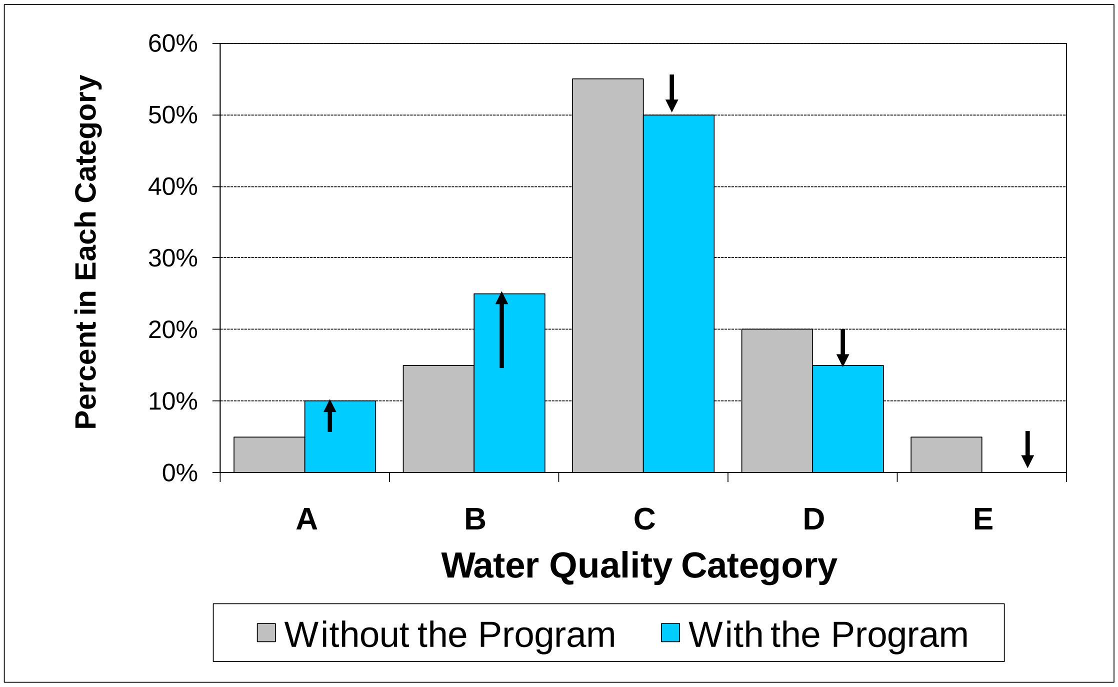
Our objective in describing the program this way was to allow the incidence of clean-up responsibility to be broadly distributed. Based on cognitive interviews and past experience, respondents are sometimes concerned about the perception of fairness or implicit property rights. With the program defined, respondents were then given textual and graphical information on how the program could be expected to improve water quality:

> *The diagram below compares projected lake conditions in HOME STATE <u>in 10 years</u>, with and without the program. The bars in grey show what lakes would be like without the program. If no action is taken to control nitrogen and phosphorus, only 20% (2 out of every 10 lakes) would be in one of the best two categories (**A** or **B**). The bars in blue show what lakes would be like with the program. 35% would be in one of the best two categories. The arrows show how the percent of lakes in the best two categories would increase, and the percent in the other categories would decrease.*

An example of visual display accompanying this text is shown below in Figure 4.

With the baseline conditions, program, and potential improvements so described, the next step was to describe how the program would affect the respondent financially, if it were to go forward. The payment vehicle, as it is referred to in the SP literature, describes how the respondent would pay for the program. Selecting an appropriate payment vehicle often poses a challenge. More specific payment vehicles make the whole scenario more realistic, but also increase the probability that respondents will vote based on their reaction to the payment vehicle rather than the scenario the researcher wants to value.

Figure 4. Graphic Used to Show Change in Water Quality



Boyle (2003) reviews payment vehicles used in different studies, including income tax increases, general increases in the cost of living, increases in utility bills, or entrance fees and donations. In the survey, we presented a general set of actions that the state would take to improve water quality. Some of these actions might be financed by taxes, while others would be paid for by individuals or businesses through utility bills or increases in the costs of other goods. After pretesting, we selected a general payment vehicle that emphasized an overall increase in their cost of living. We felt this general payment vehicle best reflected the various avenues through which respondents would have to pay for the plan we described, and in pretests respondents seemed to accept the approach. We settled on the following phrasing:

> *The changes required by the program would have a cost for all HOME STATE households. Some of the basic things people spend money on would become more expensive. For example, for homeowners, water bills or costs for maintaining septic*

systems would go up. For renters, rent or utility bills would go up. Imagine that for households like yours, starting next year, the program would permanently increase your cost of living by $V per year, or $V/12 per month.

As we describe in the next subsection, *V* is one of several bid amounts presented to the survey respondents.

The final step is the wording for the actual choice question. Again, there are different ways to phrase the choice. The NOAA panel recommends phrasing the question as a referendum. Boyle (2003) discusses the research on decision rules, including results that suggest a risk of a referendum format is that it might induce respondents to vote as "good citizens" rather than reveal their individual, self-interested WTP. Against this is research (e.g., Carson and Groves, 2007) suggesting that incentive compatible decision rules such as referenda can help minimize hypothetical bias (the tendency to answer hypothetical questions differently than if it were a real choice). We decided to use a voting referendum, including the decision rule that the program will be adopted if a majority of the voters support it. As is often done in SP studies, we included text reminding people of their budget constraint ("cheap talk"), which is designed to counteract the problem of hypothetical bias. The cheap talk reminder noted that people sometimes answer hypothetical questions differently than real questions and asked respondents to avoid this phenomenon. We also emphasized the importance of the respondent's answer for policy makers to increase the saliency of the question. After presenting the amount by which a household's costs would increase if the program were in place, respondents were given the following hypothetical choice question:

Imagine that all HOME STATE residents were allowed to vote on the program. If a majority of voters support the program, it would be implemented next year. We would *like you to think carefully about how you would actually vote in this situation. In previous research we have found that people are often more willing to vote yes when payment is only imagined than when payment is real. Therefore, we urge you to respond as though costs for your household really would go up if the program were implemented. Knowing how different HOME STATE residents would vote on this program is very important for state government decision makers. So please take time to consider both the benefits of the program and the costs to your household. Ask yourself whether you believe the lake improvement program is worth $V each year to your household, since that is less money that you would have to spend on other things. There may be good reasons for you to vote for the program and good reasons to vote against it. Only you know what is best for you and your household.*

Respondents were then asked if they would vote for or against the program. Following this initial question, respondents were asked a follow-up question soliciting if they would pay a higher or lower amount, based on their response to the first question.

3.5 Experimental Design

With the choice experiment and CV question formats established, the next step was to create an experimental design for both the choice experiment and CV sections. For the choice experiment, creating an experimental design involved determining the different values that the travel distance attribute would take, generating the set of feasible choice combinations (e.g., the universe of questions with particular attribute-level combinations we might present), and settling on choice question sets that different versions of the survey would contain. For the CV exercise we needed to determine the set of bid amounts that respondents would face, the level and amount of variability in water quality outcomes that the hypothetical program would provide, whether we would use certainty scale follow-up questions, and the extent to which our elicitation would attempt to add extra bounding information via auxiliary dichotomous choice questions. The various dimensions of the design space led us to use six survey versions that varied in the choice experiment and/or CV parameters. In what follows we describe the details of our design.

Choice Experiment

As noted in the previous section, our choice experiment used only two attributes: water quality and travel time to reach the recreation site. Thus, for purposes of experimental design, the number of values that the water quality attribute could take on was predetermined. For the four levels of the travel time attribute, we used information from our own experience, summaries of the recreation data set that we have in hand, previous studies on the length of day trips, the focus groups, and advice from our peer reviewers. Our final design included four different amounts, all expressed as one-way travel time: 20 minutes, 40 minutes, 60 minutes, and 120 minutes. With five levels for the water quality attribute and four for the travel time, our full factorial included 20 choice elements. Thus, our design space was relatively small and our experimental design task comparably simple. We decide to present each respondent with six choice tasks, and by designing six conjoint versions that varied in the composition of choice tasks, we were able to present the full factorial (absent dominant choices) to our sample. Table 3 below displays the specific choice tasks contained in each of the six versions of the survey.

Contingent Valuation

The main challenges for the CV design were determining the bid amounts and the level of water quality that the program would deliver. For the bid amounts we used previous SP research on water quality valuation (in particular Banzhaf et al., 2006) and the results from our focus groups to arrive at four annual cost levels, which would be randomly varied across

Table 3. Conjoint Experimental Design[a]

	Version 1											
Task	1		2		3		4		5		6	
Lake	1	2	3	4	5	6	7	8	9	10	11	12
Attribute 1 – Water Quality	1	3	5	3	1	4	3	4	2	3	5	4
Attribute 2 – Distance	3	2	3	4	4	1	3	2	2	1	1	4
	Version 2											
Task	1		2		3		4		5		6	
Lake	1	2	3	4	5	6	7	8	9	10	11	12
Attribute 1 – Water Quality	5	4	1	5	2	3	2	1	5	4	1	5
Attribute 2 – Distance	2	4	3	1	4	1	2	4	3	4	3	2
	Version 3											
Task	1		2		3		4		5		6	
Lake	1	2	3	4	5	6	7	8	9	10	11	12
Attribute 1 – Water Quality	5	2	1	3	3	5	2	5	1	2	4	3
Attribute 2 – Distance	1	2	4	3	4	2	4	3	3	1	1	3
	Version 4											
Task	1		2		3		4		5		6	
Lake	1	2	3	4	5	6	7	8	9	10	11	12
Attribute 1 – Water Quality	4	3	1	2	5	3	2	5	4	3	1	2
Attribute 2 – Distance	1	4	4	3	1	2	4	1	1	2	3	2
	Version 5											
Task	1		2		3		4		5		6	
Lake	1	2	3	4	5	6	7	8	9	10	11	12
Attribute 1 – Water Quality	1	3	1	4	4	2	1	4	2	3	5	4
Attribute 2 – Distance	3	1	4	3	1	4	3	2	4	2	1	2
	Version 6											
Task	1		2		3		4		5		6	
Lake	1	2	3	4	5	6	7	8	9	10	11	12
Attribute 1 – Water Quality	1	4	5	2	4	2	2	5	4	1	2	3
Attribute 2 – Distance	3	1	1	3	1	2	4	2	2	4	4	3

[a] The survey contained six SP choice tasks. In the survey, task 1 is question 16, task 2 is question 18, task 3 is question 20, task 4 is question 21a, task 5 is question 23 and task 6 is question 24a. Appendix A contains the survey instrument. Each respondent was randomly assigned one of the six versions. Attribute levels 1 to 5 for the water quality level correspond to quality levels A to E. Likewise, attribute levels 1 to 4 for travel times correspond to times in minutes of 20, 40, 60, and 120.

respondents: $24, $120, $216, and $360. We also decided to use a double bounded dichotomous choice framework. In this framework, respondents are presented with a subsequent amount that is higher or lower, depending on the initial answer. Thus, each primary bid amount has two secondary bids associated with it:

- $24: yes → $120, no → $12
- $120: yes → $216, no → $24
- $216: yes → $360, no → $120
- $360: yes → $480, no → $216.

For the water quality improvement attribute, we decided on four different levels of improvement. The survey presented a baseline distribution (see Figure 2) and an improved distribution (see Figure 3). The baseline distribution was constant for all survey respondents, and we varied the improved conditions across the four CV versions. Table 4 shows the distributions (indexed I to IV) that were presented in each version.

3.6 Pretesting and Peer Review

The survey and experimental design described above pertain to the final version of the survey that we fielded in April 2010. As part of the development process, we conducted a pretest and peer review of the survey vehicle in February and March 2010. The pretest

Table 4. Distribution of Lake Water Quality Levels for the CV Question for Baseline and Four Versions

Water Quality Index Level	Baseline	I	II	III	IV
A	5%	10%	15%	10%	20%
B	25%	25%	35%	55%	45%
C	50%	50%	40%	30%	30%
D	15%	15%	10%	5%	5%
E	5%	0%	0%	0%	0%

used 100 respondents from the same web panel used for the final survey (see more details below), each of whom completed the survey as it existed at the time. Peer reviews were conducted by Dr. Kevin Boyle of Virginia Tech University and Dr. John Whitehead of Appalachian State University. The reports submitted by Dr. Boyle and Dr. Whitehead are included as Appendix D to this document.

By and large, the pretest confirmed that our survey development strategy was effective. Our named lake lists for the recreation section were reasonably complete, in that more than 80% of trips people reported making were to lakes included in the lists. Our conjoint section included four choice tasks, and the limited item nonresponse convinced us it made sense to expand this to six choice tasks in the final survey. Analysis of the pretest conjoint data produced sensible and stable parameter estimates that suggested good scenario buy-in among our respondents. Analysis of the CV data highlighted improvements that were needed in this section. In particular, we were not able to estimate price effects precisely because a large majority of people voted "yes" for the program. Also, the pretest did not contain enough variability in quality levels to find evidence of scope in people's *WTP* for water quality. Based on these findings, we made adjustments to the experimental design for the CV section of the final survey. The peer reviewers' comments echoed our findings from the pretest. Both reviewers made small suggestions on the conjoint section of the survey, and both provided useful feedback on how we could better develop the CV section.

4. Survey Execution and Data Summary

Knowledge Networks (KN) conducted the data collection for the survey. KN maintains a web-based panel of U.S. households that were originally recruited through random-digit dialing; more recently KN has begun using address-based sampling to recruit the panel (for more information on KN, see http://www.knowledgenetworks.com). If the household does not have a computer, KN provides the household with a computer and Internet access. If the household does have a computer, KN pays for Internet access. In return, the households agree to take a specific number of surveys. KN controls the number of survey invitations panel members receive. Samples for specific surveys are drawn from the panel using probability methods.

KN sent an invitation to take the survey to 1,873 panel members age 18 or older living in our target states. The final version of the survey went to the field on April 23, 2010, and data collection was closed on May 18, 2010. In total, 1,327 individuals completed the survey, resulting in a 70.8% completion rate. The full response rate for KN surveys is much lower when panel recruitment and attrition are factored in.

Table 5 displays the sample splits arising from our experimental and state selection design.[2] Our relevant population is residents of eight southeastern stages: Alabama, Georgia, Kentucky, Mississippi, North Carolina, South Carolina, Tennessee, and Virginia. The table shows that we obtained a larger number of observations from North Carolina, South Carolina, and Virginia, which we consider the core of our study area, because it is in these states that we expect our water quality model to provide the most reliable predictions. In addition, the North Carolina Water Resources Research Institute (WRRI) funded a co-project that provided

[2] The survey data and a Stata do-file that performs all of the summaries and modeling contained in this and the following section are available as a supplement to this report.

Table 5. Sample Distribution across Survey Versions

Observations by States	AL	GA	KY	MS	NC	SC	TN	VA
	102	102	113	97	366	211	119	217
Observations by Choice Experiment Version[a]	**V1**	**V2**	**V3**	**V4**	**V5**	**V6**		
	127	135	156	127	128	137		
Observations by CV Version[b]	**I**	**II**	**III**	**IV**				
	302	353	329	343				
Observations by Water Quality Description Treatment[c]	**WQ3**	**WQ5**						
	661	666						

[a] Choice experiment versions listed in Table 3.
[b] CV versions listed in Table 4.
[c] WQ3 used three attributes to describe water quality; WQ5 used five attributes to describe water quality.

additional resources for the North Carolina component of the sample. As described above, our choice experiment design included six versions of six choice tasks each. Sixty-one percent of the respondents were recreators or likely recreators, which provided 810 respondents in this subsample. Those who did complete this section were split approximately evenly among the six survey versions. The CV section used four different water quality levels, so this part of the survey had four variations. These are also approximately evenly divided among the full sample. Finally, as discussed in the previous section, we included a split sample design in the survey, which simplified the water quality communication section for half of the sample. Approximately half of the respondents who completed the choice experiment received descriptions of our water quality rating system that excluded the odor and algae attributes, while the remaining 666 people received the full description.

4.1 Sample Characteristics

Table 6 contains basic summary statistics for several of the socioeconomic variables available in our data set. Because the KN panel is designed to be nationally rather than state representative, these summaries are unlikely to match the distributions in individual states

Table 6. Summary Statistics for Socioeconomic Variables

Variable	Mean	Std. Dev.	Median	Min	Max
Household income ($1,000s)	57.04	40.07	50	5	175
Household size	2.61	1.41	2	1	10
Respondent age	48.76	15.88	50	18	94
Driving distance to nearest lake	38.59	36.66	30	0	240
Home-owning household	0.71	—	—	0	1
Respondent full-time work	0.56	—	—	0	1
Respondent retired	0.18	—	—	0	1
Respondent male	0.44	—	—	0	1
Respondent high school	0.26	—	—	0	1
Respondent some college	0.32	—	—	0	1
Respondent bachelor degree or higher	0.33	—	—	0	1
Respondent Hispanic or nonwhite	0.265	—	—	0	1

exactly. For example, according to U.S. Census figures the median household income in North Carolina is $46,500 annually, while the median for our survey respondents in North Carolina is $50,000. Likewise, 36% of our North Carolina sample respondents have at least a 4-year college degree, while the corresponding U.S. Census figure is 22%.

These figures suggest our sample is comparatively wealthier and more educated than the overall population in the states included in our sample. For the objectives of our study related to water recreation this is less of a concern than it may seem, because past research has shown that basic recreation behavior (as opposed to specialized/exotic activities) is not substantially influenced by income and education (see Phaneuf and Smith, 2005, for a discussion of income effects in recreation models).

4.2 Behavioral Summaries

Among our 1,327 respondents, 427 (32%) reported having made a recreation day trip (i.e., without an overnight stay) to a lake in the previous 12 months. The median person in this group visited two different lakes during the 12 months prior to the survey and completed four trips.

As is typical for recreation data, the distribution of trips is skewed by avid users so that the average number of trips per individual is 9.4. We solicited detailed information from the trip-taking individuals on their activities, group composition, and location choices. Table 7 contains a summary of the main activities that respondents participated in and a summary of their additional/auxiliary activities. The final row of Table 7 lists the percent of respondents who selected each activity for the SP questions. Activity choices are relevant in that the degree of contact with the water may condition people's attitudes toward water quality. Almost a third of the trip-taking people in our survey reported they went swimming during at least one of their trips. Swimming, fishing, and nature viewing were the most frequently reported main activities, through the table suggests there is considerable heterogeneity in activities. Nearly half of the respondents used trails near lakes for walking or running. Respondents generally visited lakes for recreation in the company of others. Only 6% of people completed a typical trip alone, while the remainder were equally split among those whose group contained only other adults and those whose group included a mix of adults and children.

Although our study focuses primarily on day trips, we did ask people to report the degree to which they participated in overnight visits to lakes. Approximately 18% of people in our

Table 7. Activity Summaries for Actual Trips and Activity Selected for SP Questions

	Swimming	Fishing	Motor Boating	Non-motor Boating	Nature Viewing	Organized Event	Running or Walking
Main activity	14%	23%	11%	4%	17%	10%	17%
Additional activity	33%	36%	19%	9%	67%	25%	46%
Activity selected for SP questions	16%	23%	11%	4%	15%	10%	18%

sample took at least one overnight trip to a lake. Among them, the median number of trips is 2 and the average is 4.35. The latter is heavily skewed by a few avid trip takers who reported overnight trips in excess of 40 visits.

4.3 Water Quality Beliefs Summary

Table 8 presents a summary of how respondents in different states rated the water quality at lakes in their states. This table uses responses to question 11 (see Appendix A), which asked people to indicate the quality level they thought was most common for lakes in their state. Most of the answers clustered around levels B and C. There are some differences across the states. For example, 72% of people in Alabama thought lakes in their state fell in the A or B (the two best) range, while only 51% of people in Tennessee reported similar beliefs. In general, very few people thought lakes in their state fell into the worst category (level E).

Table 8. Summary of Beliefs about Lake Water Quality by State

	Water Quality Category					
State	**A**	**B**	**C**	**D**	**E**	**Sample Size**
AL	17%	54%	22%	5%	2%	102
GA	17%	38%	33%	8%	4%	102
KY	10%	52%	27%	8%	1%	113
MS	13%	43%	29%	8%	4%	97
NC	10%	47%	30%	9%	2%	366
SC	16%	44%	28%	9%	0%	211
TN	6%	45%	30%	8%	7%	119
VA	13%	47%	27%	10%	1%	217
All states	12%	46%	29%	9%	2%	1,327

5. Basic Analysis

In this section we present the results from several analyses using the conjoint and CV data. For each data type our strategy is to first present models that help us understand the variability in the data, provide some initial sense of water quality valuations, and allow us to make decisions on what our preferred policy-relevant models will look like. We then present what we consider our policy-relevant results, which we use for our case study in the following section.

5.1 Choice Experiment Analysis

We begin by looking at the conjoint data in detail. Recall that there were 810 respondents who qualified for the conjoint exercise – 427 people reported that they visited a lake for a day trip in the past 12 months and an additional 383 indicated they would likely visit a lake in the next 12 months. Each of the 810 respondents received six conjoint questions; some people did not complete all six questions, so our final sample size consists of 4,849 observations. The analysis in this subsection is based on these observations.

Recall that our choice experiment questions proceeded in two steps. First, people were asked to compare two lakes that differed only in the distance from home and water quality level. Second, they were asked to indicate if they would actually make a trip, if these were the only two options. We used this two-step approach to maximize the information gathered on preferences per question, while maintaining the realism implied by an opt-out choice. Among the 4,849 choices, 26% selected the “no trip” option (or the opt-out). Recall as well that people were asked to name the activity and group composition that would define their trip and thereby condition their answers to the first four choice tasks. The final row in Table 7 shows the distribution of activity choices people selected for the choice experiment. Thirty-nine percent of people selected the activities fishing or swimming that involve contact with the water, while 33% selected near-shore activities such as nature viewing or running/walking. In terms of group composition, 4% reported they would make the trip alone, 46% said they would be in the company of other adults, and 50% said they would be in a group that included children.

The economic model typically used to analyze choice experiment data of the type we collected is the random utility maximization (RUM) model. The underlying assumption in this model is that people make choices to maximize their utility (i.e., their well-being). In our choice tasks, we assume that respondents select the option that they believe would provide them with the highest satisfaction. We model this by specifying a respondent's conditional indirect utility function as

$$U_{ict} = V_{ict} + \varepsilon_{ict}, \quad i = 1, \ldots, I, \quad t = 1, \ldots, T, \quad c = 0, 1, 2, \tag{1}$$

where the utility available to respondent i on choice task t from selecting option c consists of two parts: an observable component V_{ict} and an unobservable component ε_{ict}. The former is a function of measured covariates and parameters to be estimated, such as the level of water quality and travel time for options 1 and 2 (the lake options), while the latter captures the component of the respondent's tastes that is not reflected by any measured variables. For example, in our application the baseline model is

$$\begin{aligned} V_{ict} &= \beta_1 \times time_{ict} + \beta_2 \times qual_{ict}, \quad c = 1, 2 \\ V_{ict} &= \beta_0, \quad c = 0, \end{aligned} \tag{2}$$

where $time_{ict}$ and $qual_{ict}$ are the values of the travel time and water quality attributes from the conjoint design, and (β_0, β_1, β_2) are utility function parameters to be estimated. Note that $qual_{ict}$ is written as a continuous variable, although the actual quality levels in our survey are discrete. In our initial models we use the cardinal progression of water quality levels (A, B, , C, D, E) to code $qual_{ict}$ = 1 if lake option *c* had water quality A, $qual_{ict}$ = 2 if lake option *c* had water quality B, and so on out to $qual_{ict}$ = 5 if the lake option had water quality E. We subsequently explore models that use dummy variables for each of these discrete levels.

The operational assumption in RUM models is that the analyst knows the distribution from which ε_{ict} is drawn but does not know its exact value. Thus, before observing a choice, the analyst can only know the probability of seeing a particular outcome, conditional on parameter values. Because option *c* will be selected if its utility value is highest (i.e., if $U_{ict} \geq U_{ist}$ for all $s \neq c$), we can derive this probability (Pr) as

$$\begin{aligned} \Pr_{it}(c) &= \Pr\left[U_{ict} \geq U_{ist}, \forall s \neq c\right] \\ &= \Pr\left[V_{ict} + \varepsilon_{ict} \geq V_{ist} + \varepsilon_{ist}, \forall s \neq c\right] \\ &= \Pr\left[V_{ict} - V_{ist} \geq \varepsilon_{ist} - \varepsilon_{ict}, \forall s \neq c\right] \\ &= \Pr\left[\varepsilon_{ist} - \varepsilon_{ict} \leq V_{ict} - V_{ist}, \forall s \neq c\right]. \end{aligned} \tag{3}$$

Equation (3) shows that with knowledge of the distribution for ε_{ict}, we can write an expression for the probability of observing any choice outcome as a function of the covariates and parameters contained in V_{ict}. By matching these *ex ante* probability expressions to the *ex post* responses, we can estimate the parameters of the utility function by maximum likelihood, where the log-likelihood function is

$$LL(\beta) = \sum_{i=1}^{I} \sum_{t=1}^{T} \sum_{c=0}^{2} d_{ict} \ln \Pr_{ict}, \tag{4}$$

and d_{ict} in an indicator variables equal to one if option *c* was selected on choice task *t*, and zero otherwise. In writing equation (4), we have for simplicity assumed the *T* choices made by respondent *i* are independent, and in this section we do not exploit the additional information that is provided by the two-step choice solicitation. Specifically, we do not use the ranking information provided by the initial choice between the two designed options when the opt-out option is ultimately selected; instead we treat the outcome as arising from a single trinomial choice. We do this for simplicity and transparency in this section, and incorporate the richer ranking information when we present our final policy models. Nonetheless it is necessary to cluster the standard errors of estimates at the level of the respondent, which we do in all our model runs. If we assume that ε_{ict} is distributed type I Extreme Value, the familiar conditional logit model arises.

Estimates of β_0, β_1, and β_2 from equation (2) are useful for measuring the trade-offs people are willing to make between travel time and water quality. To see this, it is helpful to distinguish between marginal utilities and marginal values. In equation (2) β_1 is the marginal disutility of travel time, which implies that $-\beta_1$ is the marginal utility of time itself. The parameter β_2 is the marginal disutility utility of a one-unit increase in the water quality variable (recall that lower values correspond to better quality). Although the signs of these parameters do have qualitative meaning, they do not have a quantitative interpretation in isolation. However, we can use the two together to compute the marginal value of a change in water quality, expressed relative to time, by totally differentiating the conditional utility function with respect to *time* and *water*:

$$\Delta U_{ict} = \beta_1 \Delta time_{ict} + \beta_2 \Delta qual_{ict}. \tag{5}$$

If we hold utility constant at a reference level so that $\Delta U_{ict} = 0$, then

$$\frac{\Delta qual_{ict}}{\Delta time_{ict}} = -\frac{\beta_2}{\beta_1}. \tag{6}$$

Conditional on making a trip, this ratio tells us the rate at which people are willing to trade travel time for water quality, while holding fixed a reference level of well-being. Equation (6) is useful because it allows us to express the value of an improvement in the water quality index at a lake in terms of the extra travel time a person would accept to have the improvement.

Table 9. Conditional Logit Results by State, Parameter Estimates with Z-Statistics

	All	AL	GA	KY	MS	NC	SC	TN	VA
Opt-out	−3.855	−3.342	−3.519	−3.808	−3.742	−4.031	−3.953	−3.757	−4.113
	(−31.81)	(−7.15)	(−8.05)	(−9.86)	(−7.56)	(−17.34)	(−13.15)	(−9.53)	(−15.21)
Travel time	−0.008	−0.007	−0.008	−0.008	−0.008	−0.009	−0.010	−0.009	−0.009
	(−20.05)	(−4.66)	(−5.16)	(−5.98)	(−4.44)	(−10.34)	(−9.4)	(−6.55)	(−8.57)
Water quality	−0.824	−0.747	−0.736	−0.782	−0.784	−0.872	−0.829	−0.795	−0.899
	(−30.36)	(−6.75)	(−7.39)	(−9.12)	(−7.35)	(−17.78)	(−12.48)	(−8.99)	(−13.03)
Number of choices	4,849	390	396	391	317	1,389	816	448	702
Marginal value water quality improvement in hours (standard errors)[a]	1.625 (0.08)	1.807 (0.36)	1.607 (0.28)	1.693 (0.31)	1.584 (0.32)	1.694 (0.15)	1.402 (0.14)	1.516 (0.22)	1.753 (0.20)

[a] Standard errors computed using the delta method (Greene 2000).

In what follows we discuss estimates from several different models and subsets of the respondents, using the conditional logit assumption in all cases. We begin with the simple specification in equation (2). Table 9 contains estimates from this model for the entire sample, as well as estimates obtained for subsamples corresponding to each state in our study region. The table shows that the coefficient estimates are significant and intuitively signed across all the models. The negative coefficient on *Opt-out* shows that respondents on average found it preferable to select a trip option, rather than not participate in recreation on a given choice occasion. Since we asked people to image they were planning a trip this means the designed alternatives on average met minimum quality standard thresholds. The estimates also show that a lake site is less attractive if it is further away, or if its water quality index is higher. Said another way, the utility of a site can be made greater if the value of its water quality index is reduced (water quality is improved). In the full sample, the ratio of the water quality parameter to the travel time parameter ($-\beta_2/\beta_1$) is 1.625, which suggests that people value a one-unit improvement in the value of the quality index (e.g., from *qual* = 3 to *qual* = 2) the same way that

Table 10. Conditional Logit Results by Main Activity and Type of Group, Parameter Estimates with Z-Statistics

	Swim	Fish	Boating[a]	Walk	Nature Viewing	Trip with Children
Opt−out	−4.022	−4.222	−4.369	−3.599	−4.012	−3.787
	(−13.68)	(−15.53)	(−13.88)	(−12.51)	(−14.46)	(−22.75)
Travel time	−0.008	−0.008	−0.010	−0.010	−0.009	−0.008
	(−7.45)	(−9.52)	(−8.07)	(−9.52)	(−9.4)	(−14.29)
Water quality	−0.897	−0.980	−0.831	−0.735	−0.828	−0.811
	(−14.58)	(−14.31)	(−13.18)	(−12.33)	(−13.86)	(−21.17)
Number of choices	723	1026	658	858	916	2404
Marginal value water quality improvement in hours (std. error)[b]	1.876 (0.22)	2.013 (0.20)	1.364 (0.15)	1.287 (0.12)	1.557 (0.15)	1.699 (0.11)

[a] Includes both motor boating and nonmotor boating.
[b] Standard errors computed using the delta method (Greene 2000).

they value 1.625 hours of extra time. When we split the sample for the different states, we see that marginal values for water quality held by residents across the study region vary from a low of 1.40 hours for South Carolina to a high of 1.80 for Alabama. The standard errors on the state-level estimates generally suggest most are statistically equal to the full sample model at conventional significance levels.

Table 10 provides estimates by activity and group composition. These data allow us to examine the impact of activity (and whether the activity involves contact with the water) on the value of water quality. We ran separate models for the different activities and for those who said they would travel with children. Our estimates generally suggest that people who participate in swimming or fishing have a higher marginal value for water quality compared with those who participate in boating, walking, or nature viewing. Thus, there is some evidence of heterogeneity in water quality values arising from people’s activity preferences. Comparing the first column in Table 9 with the subsample who said they would travel with children, the results suggest that people whose group includes children do not seem to value water quality differently from the full sample.

Although these results provide insights into the data, they do not allow us to express the value of water quality in monetary terms at the level of the individual. For this we need to convert the time cost of travel a person must bear to reach the site into a monetary equivalent. A common way of computing the value of time (i.e., the opportunity cost of time) in environmental economics is to use a fraction of the average wage rate (see Phaneuf and Smith, 2005). A typical choice is to use 0.33 of the wage rate, so we computed the opportunity cost of time using oct_i = 0.33 × $income_i$/2,000, where 2,000 is the approximate number of work hours in a year (Ceserio, 1976). With this we create a new variable

$$price_{ict} = oct_i \times \frac{time_{ict}}{60} + \left(gas / mpg + dep\right) \times \frac{time_{ict}}{60} \times speed, \quad c = 1, 2, \tag{7}$$

where *gas* = 2.75 is the per-gallon price of gasoline, *mpg* = 20 is our average miles per gallon assumption, *dep* =0.20 is our assumption for vehicle depreciation per mile, we assume an average speed of 45 miles per hour to translate travel time into out-of-pocket travel costs. The time variable is divided by 60 to change the units of time into hours. We replace $time_{ict}$ in equation (2) with the new variable $price_{ict}$. The variables *gas*, *mpg*, *dep*, and *speed* are set to a single value for all respondents, the opportunity cost of time (*oct*) varies by respondent, and travel time (*time*) varies across respondents and alternatives. The transformation in (7) there will therefor affect parameter estimates.

The first column of Table 11 contains estimates for a simple model including only price and water quality. As with the simpler model with time as a covariate, both coefficient estimates are negative, so the qualitative interpretation from earlier is still valid. However, for this model the ratio of coefficients provides an estimate of the marginal value in dollars of a unit change in

Table 11. Conditional Logit Estimates Using Variable Price under Different Specifications with Z-Statistics

	Model I	Model II	Model III	Model IV
Opt out	−3.491	−3.498	−3.981	—
	(−27.2)	(−27.23)	(−25.01)	
Price	−0.017	−0.017	−0.017	−0.018
	(−15.18)	(−15.18)	(−14.93)	(−14.67)
Water	−0.765	−0.804	−1.173	—
	(−27.45)	(−25.57)	(−14.86)	
Water×WQ3	—	0.074	—	—
		(2.40)		
water squared	—	—	0.069	—
			(5.47)	
Qual A (δ_A)	—	—	—	2.816
				(24.94)
Qual B (δ_B)	—	—	—	2.071
				(19.75)
Qual C (δ_C)	—	—	—	1.063
				(12.34)
Qual D (δ_D)	—	—	—	0.208
				(2.33)
Qual E (δ_E)	—	—	—	−0.033
				(−0.37)

Marginal Value of Water Quality (dollars)

	Model I	Model II	Model II	Model III	Model III	Model IV
	All	**WQ5[a]**	**WQ3[a]**	**Quality = B**	**Quality = C**	**Change from C to B**
Marginal value water quality in $ (std. error)[b]	$44.84 (2.62)	$47.05 (2.84)	$42.71 (2.69)	$51.68 (3.16)	$31.73 (3.15)	$57.23 (4.43)

[a] WQ5 indicates respondents who received the survey version with five characteristics describing water quality. WQ3 denotes those who received the version with three characteristics.
[b] Standard errors computed using the delta method (Greene 2000).

water quality. We find that people are, on average, willing to pay almost $45 per lake visit for a one-unit improvement in water quality (i.e. to move to the next higher water quality category).

The additional models in Table 11 explore different ways of including the water quality variables in the specification, with an emphasis on exploring nonlinearity. Model II tests whether people who received the simpler description of the water quality with only three attributes respond differently to the quality level. The coefficient on the interaction term between water quality and a dummy variable indicating the person received the simpler water quality description (WQ3) is positive and significantly different from zero, suggesting there is a difference in response. People who received the simpler description of water quality placed a lower value on water quality improvements, though the estimates for marginal WTP suggest preferences are qualitatively similar. Model III explores the extent of nonlinearity in how people respond to water quality. We find a significant and positive coefficient on the *water squared* term, which implies that the marginal WTP for water quality increases as water quality improves. For example, the marginal WTP for a change when the baseline quality level is *water quality* = 2 (level B) is $51.68, and it falls to $31.73 when *water quality* = 3 (level C). This provides evidence that there is a nonlinear relationship between the level of water quality and the marginal WTP for a change in quality, though the curvature suggested by the coefficient estimates is counter intuitive. As we show below using a more flexible specification, this finding is an artifact of the quadratic functional form and not a perverse aspect of our data.

Model IV moves away from using the five levels of water quality in the design as a continuous variable and instead includes a dummy variable for the discrete quality levels so that

$$\begin{aligned} V_{ict} &= \beta_1 \times price_{ict} + \sum_{k=A}^{E} \delta_k d_{ict}^k, \quad c = 1,2 \\ V_{ict} &= 0, \quad c = 0, \end{aligned} \tag{8}$$

where d^k_{ict} = 1 if lake c has quality level k, for k = A, B, C, D, E, and zero otherwise, and δ_k is a quality level specific coefficient that will be estimated. Equation (8) is the most flexible way that we can characterize preferences for the different water quality levels, in that it allows the WTP for a change in the water quality index to increase or decrease throughout its range. Note that we have included dummy variables for all five quality levels for ease of interpretation, meaning we have normalized the constant term to zero for c=0. In equation (8) we expect $\delta_A > \delta_B \ldots > \delta_E$, because water quality level A is the highest and E is the lowest. The parameter values in Table 11 follow this pattern, meaning that our estimates pass a scope test. More specifically, respondents have a clear preference ordering for better water quality; conditional on travel cost people prefer the destination with better water quality. Sensitivity to scope has emerged in the stated preference literature as an important indicator of validity (see Kling et al. 2012 for discussion). As such model IV provides strong evidence of the quality of our stated preference data. The estimates from this model imply the WTP predictions (standard error) for a one-unit change in quality starting from different baseline changes are

- Level E to Level D: $13.66 (4.68)
- Level D to level C: $48.50 (5.16)
- Level C to Level B: $57.22 (4.43)
- Level B to Level A: $42.25 (4.08)

The biggest gains are for movements from D to C and C to B. These results suggest a nonlinear relationship based on the notion that the gains from marginal improvements are small when water quality is quite low or quite high, relative to improvements that shift water quality from levels that limit activities (e.g., E or D) to levels that enable them (C and B).

These preliminary findings are suggestive of several trends in our choice experiment data. First, people are willing to trade off travel time for better water quality when making recreation destination decisions. It seems that improvements in water quality do have value and the value is captured by our model. Second, water quality values are likely to be heterogeneous across activity types; water-contact trips such as fishing and swimming are associated with higher values than near-shore experiences, such as using trails. Third, group composition does not seem to matter much for water quality values. Finally, the role that water quality plays in people's preference for recreation sites is nonlinear. We use these insights in constructing our policy-relevant model.

Policy Choice Experiment Model

Recall that the output from the models described in Phaneuf et al. (2009) consists of predictions for the discrete probability distribution of a lake's (or a collection of lakes') quality levels. Given this, it is useful to use estimates from the choice experiment models that include dummy variables for each of the discrete water quality levels. In addition, these estimates provide the most flexible representation of the nonlinearity in peoples' preferences. Table 12 provides estimates for two types of models: a single column that includes all the observations drawn for the study, and several columns that are state specific in that they only include observations drawn from the particular state. The parameter estimates shown in the "All" column in Table 12 differ slightly from those for the comparable model shown in Table 11 because we have now used the extra information that our two-step opt-out option provides. In particular, if a person selects lake 1 over lake 2 but then indicates she would not take a trip if the two designed lakes were her only options, we know that utility(no trip)>utility(lake 1)> utility(lake 2). We have used this extra ranking information in the construction of the likelihood function for all the models presented in Table 12, in order to maximize the information content that goes into constructing our policy-use models. The tradeoff is a slightly more complicated and less transparent likelihood function (Stata code for this model in included in the appendix).

We suggest using the models in Table 12 for policy purposes given that they are relatively parsimonious. Though it is possible to include more accounting for observable

Table 12. Conditional Logit Policy Estimates with Z-Statistics for Choice Experiment Data

Variable/ Parameter	All	AL	GA	KY	MS	NC	SC	TN	VA
Price (β_1)	**−0.016**	−0.015	−0.017	−0.011	−0.012	−0.016	−0.017	−0.018	−0.016
	(−16.19)	(−3.82)	(−3.06)	(−3.39)	(−3.28)	(−8.86)	(−7.27)	(−7.01)	(−7.24)
Qual A (δ_1)	**2.560**	2.204	2.472	2.409	2.081	2.716	2.514	2.625	2.954
	(24.82)	(5.63)	(6.56)	(7.20)	(5.40)	(13.50)	(9.70)	(8.28)	(11.91)
Qual B (δ_2)	**1.938**	1.749	1.915	1.501	1.667	2.088	1.946	2.084	2.109
	(20.56)	(4.95)	(5.40)	(4.19)	(4.69)	(11.98)	(8.38)	(7.13)	(8.96)
Qual C (δ_3)	**0.983**	0.811	1.173	0.886	1.020	0.988	1.022	0.898	1.060
	(12.48)	(2.83)	(4.67)	(2.89)	(3.57)	(6.38)	(5.38)	(3.39)	(5.70)
Qual D (δ_4)	**0.176**	−0.108	0.252	0.127	−0.015	0.249	0.231	0.141	0.259
	(2.21)	(−0.37)	(0.91)	(0.42)	(−0.05)	(1.63)	(1.18)	(0.52)	(1.36)
Qual E (δ_5)	**−0.109**	−0.097	0.002	0.025	−0.247	−0.830	−0.123	−0.052	−0.205
	(−1.37)	(−0.36)	(0.01)	(0.09)	(−0.68)	(−0.40)	(−0.63)	(0.18)	(−0.96)

heterogeneity (e.g., allowing the marginal utilities for water quality to differ with activity and group composition), our sense is that the specific recreation data needed to support policy modeling via our protocol typically do not allow analysts to divide aggregate lake visitation out by activity or group composition. Thus, it seems better to assess the average per-trip WTP using models that do not include heterogeneity. We stress that the model estimates in Table 12 are not biased by their absence of interaction effects – they simply fold any and all heterogeneity into a single (average) estimate.

We have provided state-specific estimates to allow for the possibility of using spatially explicit state-level estimates if this is deemed important. We stress, however, that the limited number of observations in any given state (aside from North Carolina, where additional resources were available to increase the state sample size) means there is a cost in lost precision of using the state-specific estimates. For this reason in what follows we focus our attention on the average estimates obtained using the full sample. The per-trip WTP estimates (standard error) of discrete one-unit changes in water quality are summarized as follows:

- Level E to Level D: $18.38 (4.34)
- Level D to level C: $52.05 (4.94)
- Level C to Level B: $61.55 (4.30)
- Level B to Level A: $40.16 (4.18)

The small differences from the similar estimates presented above arise due to the slightly different coefficient estimates arising from the full information likelihood function. While the point estimates differ somewhat, the standard errors suggest the differences are not statistically significant. As noted above, we view these are the preferred estimates given their full use of the information content in the data.

In our policy setting we will generally only know the probability that a lake is in a particular category. Computing per-trip WTP estimates in this case using the results in Table 12 is computationally straightforward. Denote the probability that the policy lake has baseline water quality level k by p_k, for k = A, B, … , E. Furthermore denote the probability that the policy lake has counterfactual water quality level k by p_k^c for k = A, B, … , E. The expected per-trip WTP for the counterfactual quality improvement is

$$E(WTP) = \frac{1}{-\beta_1}\left[(p_A^c - p_A)\delta_1 + (p_B^c - p_B)\delta_2 + \ldots + (p_E^c - p_E)\delta_5\right], \tag{9}$$

where the parameter values are taken from the “All” column in Table 12. With the choice experiment estimates and predictions from the eutrophication production function, a mapping from a change in lake-level assessed water quality to the per-trip recreation benefits of the improvement is available. In Section 6 we provide a simple case study on using the models together.

Table 13. Percent Voting Yes by Cost in CV Question

	$24	$120	$216	$360
Percentage voting yes	73%	60%	50%	42%

5.2 Contingent Valuation Analysis

We now turn our attention to the CV data. Across the entire sample of 1,327 and all bid levels, 56% of people responded positively to the referendum question. Breaking this out by the four specific bid amounts as shown in Table 13 confirms that people were responsive to costs. In stated preference studies it is generally considered an indicator of validity when quality improvement programs are less attractive to respondents when they are more costly (see once again Kling et al. 2012). Thus the responsiveness to cost provides an additional indicator of the quality of our SP survey.

Breaking the response summary out by the four program levels is also useful. Recall that the distribution of water quality in the four scenarios was such that I<II<III<IV.[3] If there is sensitivity to scope, we would expect the proportion of yes responses to increase as we move from I to IV. Broadly speaking, we see this type of pattern in the responses, but the magnitudes are not as clear-cut as for the cost variable. We find that 55% of those who received version I or version II responded yes, whereas 58% responded yes for versions III and IV. We examine this concept of scope in more detail below.

Our survey asked people to report how certain they were about their answer to the referendum question. Forty percent reported they were “very certain,” 50% “somewhat certain”, and 10% “not certain at all”. Among the people who were uncertain about their answer, only 41% answered “yes” to the referendum, while among those who were very certain 58% answer

[3] The difference between III and IV is in the percent of lakes in the top two water quality categories.

“yes”. These figures suggest uncertain respondents tended to vote “no” on the referendum, meaning our sample did not engage in the “yea saying” that has been identified as a potential threat to validity. Several studies have looked at the relationship between the respondent’s self-reported degree of certainty about their answer and the potential for hypothetical bias (the difference between responses to a hypothetical scenario and a real choice). Champ et al. (1997) and Blumenschein et al. (2008) both found more evidence of hypothetical bias among uncertain respondents. Respondents who were certain of their responses showed little or no evidence of hypothetical bias. In the results presented below (see Table 14), we compare the results using the respondents’ original votes and certainty-adjusted votes where “yes” votes by respondents who indicated they were “not certain at all” are recoded as “no” votes.

In the remainder of this subsection we consider parametric models using the CV data. Each person answered one CV question based on one of the quality change treatments shown in Table 4. The sample was divided approximately evenly across the four treatments. To explore these data we first look at models that examine the four program levels discretely, and then models that specify water quality as a continuous variable. The econometric structure is based on a utility difference (ΔU) framework. In answering the CV question, the model assumes people choose to vote for or against the program based on whether the program (including its annual cost) provides an increase or decrease in utility compared to conditions without the program. Given this, our baseline model is

$$\Delta U_i = \Delta V_i + \varepsilon_i = \gamma_1 bid_i + \alpha + \sum_{j=II}^{IV} \delta_j Z_{ij} + \beta X_i + \varepsilon_i, \quad i = 1, ..., N, \tag{10}$$

where bid_i is the annual cost for the program presented to respondent i, and the indicator variable Z_{ij} takes the value one if the respondent answered the survey version with program j, for $j = II, III, or IV$, and zero otherwise (program I is the omitted category, the effect of which is captured by the intercept term). The variable X_i represents a vector of individual characteristics that may also affect the change in utility. Together, these three components make up the systematic portion of the change in utility (ΔV). If we assume that the random component ε_i follows a standard logistic distribution, the probability of voting for the program is

$$\Pr(yes_i) = \left(1 + \exp(-\Delta V_i)\right)^{-1}, \tag{11}$$

which implies that the expected willingness to pay for program j by household i is

$$E(WTP_{ij}) = -\left(\alpha + \delta_j + \beta X_i\right) / \gamma_1. \tag{12}$$

These formulas arise from the logistic error and form of utility difference function, respectively. In particular, the term in parenthesis on the right side of (12) is the gross utility improvement generated by the program; scaling this by the marginal utility of income ($-\gamma_1$) converts this into the dollar equivalent of the utility change, which is the maximum willingness to pay to have the change. The left hand side is an expectation since the formula does not include the (zero mean) random variable ε_i.

An alternative way of modeling the role of water quality is to use a continuous index defined as

$$qual_j = (1\times p_j^A) + (2\times p_j^B) + (3\times p_j^C) + (4\times p_j^D) + (5\times p_j^E), \tag{13}$$

where p_j^q is the percentage of lakes in water quality category q (q=A,B,C,D,E) under scenario j, where j=0 is the baseline and j=I,II,III,IV represents the four designed scenarios. According to our design the baseline index value is $qual_0$=3.05, while the improved index values are $qual_I$=2.70, $qual_{II}$=2.45, $qual_{III}$=2.30, and $qual_{IV}$=2.20. Using the continuous quality index our utility difference model becomes

$$\Delta U_i = \gamma_1 bid_i + \delta_1 \ln(\Delta qual_i + 1) + \delta_2 \ln(\Delta qual_i + 1)\times X_i + \alpha + \beta X_i + \varepsilon_i, \tag{14}$$

where $\Delta qual_i = qual_0 - qual_j$ is the improved quality level presented to respondent *i* and we have used a log transformation. For this model the expected WTP for a specific change in quality is

$$E(WTP_i) = -\left[\delta_1 \ln(\Delta qual + 1) + \delta_2 \ln(\Delta qual + 1) \times X_i + \gamma + \beta X_i\right] / \gamma_1. \tag{15}$$

Table 14 presents coefficient estimates for six different dichotomous choice logit models. The first three columns correspond to the program dummy variable specification in equation (10) and the last three columns correspond to the continuous quality difference specification in (14). For each specification we present sensitivity analyses that show how the estimates change when we control in different ways for responses reported to have been "very uncertain". We compare results for three different ways of coding the dependent variable. In columns 1 and 4 we use respondents' original votes (labeled *vote*) without adjustment. In columns 2 and 5 we use a certainty-adjusted vote (labeled *vote recode*) in which respondents who indicted "not certain at all" where coded as "no" votes, regardless of the actual vote. In columns 3 and 6 we drop responses that indicated "not certain at all" (labeled *vote certain*), so that our analysis includes N=1,182 for these models.

In each of the models the coefficient on the bid level is negative and statistically significant at the $p<0.01$ level, confirming that higher costs reduce the utility of the program and the likelihood of a yes vote (as suggested by the summary statistics, this is an indicator of validity). The first three columns show that income is not a statistically significant determinant of people's vote; based on this and other statistical tests the income variable (including interactions) was dropped from the later three specifications. We also examined the effects of several other respondent- and household-specific characteristics on preferences for the program. We find that those who have used or expect to use lakes for recreation and those with post-secondary education are statistically more likely to vote in favor of the program. Other

Table 14. Logit Regression Analysis of CV Survey Responses (standard errors in parenthesis)

	Vote (1)	Vote recoded (2)	Vote certain (3)	Vote (4)	Vote recoded (5)	Vote certain (6)
Bid	-0.00391*** (0.000480)	-0.00408*** (0.000484)	-0.00423*** (0.000515)	-0.00355*** (0.000460)	-0.00389*** (0.000465)	-0.00390*** (0.000494)
Program II	0.0239 (0.164)	0.0107 (0.164)	0.0232 (0.175)			
Program III	0.202 (0.168)	0.297* (0.168)	0.326* (0.180)			
Program IV	0.176 (0.165)	0.267 (0.166)	0.300* (0.177)			
ln(Δ*qual*+1)				1.038*** (0.250)	0.636** (0.248)	1.295*** (0.272)
Income	-0.0126 (0.0136)	-0.00676 (0.0136)	-0.0143 (0.0143)			
College	0.290** (0.130)	0.343*** (0.130)	0.326** (0.138)			
triplastyr	0.596*** (0.139)	0.773*** (0.139)	0.642*** (0.149)			
tripnextyr	0.401*** (0.141)	0.531*** (0.141)	0.377** (0.151)			
ln(Δ*qual*+1)×college				0.500** (0.244)	0.601** (0.243)	0.547** (0.261)
ln(Δ*qual*+1)×trplastyr				1.197*** (0.276)	1.495*** (0.276)	1.256*** (0.297)
ln(Δ*qual*+1)×tipnextyr				0.785*** (0.278)	1.006*** (0.279)	0.714** (0.297)
constant	0.560*** (0.182)	0.219 (0.182)	0.612*** (0.197)			
Observations	1318	1318	1182	1318	1318	1182

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

characteristics, such as age, sex, race, and marital status were found to be individually and jointly statistically insignificant and were therefore excluded from the models that we present.

By varying the water quality program descriptions across respondents we are able to examine how differences in the size of water quality improvements affect responses. In columns 1 through 3 the water quality improvements are represented as program dummy variables, where program I is the omitted category in the regressions. Because water quality outcomes are better as we progress from program I to program IV the parameters on the dummy variables represent incremental increases in the utility of a yes vote. Although the estimates have the expected positive sign, we find statistical significance for only one parameter in the *vote recoded* model and two parameters in the *vote certain* model. We conclude from these positive estimates that scope effects are likely present, but that the variability in quality levels amongst the programs is too small to detect differences at the level of flexibility implied by the dummy variable model. While our design could have varied the differences among programs to a larger degree, and thereby increased the power to identify scope effects without functional form assumptions, we were constrained by the need to maintain credibility in the size of the programs' deviations from the baseline.

Given this our last three sets of estimates use the specification in equation (6), where the log transformation of the continuous quality attribute imposes a smooth diminishing marginal utility of the quality change. Columns 4 through 6 restrict the constant term and the level effects of the respondent characteristics to zero, because joint tests of these restrictions could not be rejected at the 0.10 significance level. An advantage of this outcome is that it constraints the utility change (and by extension, willingness to pay) to be zero when $\Delta qual=0$, as would be expected. In all three models the size of the water quality improvement has a positive and statistically significant effect ($p<0.01$) on the utility difference. In addition, the interaction terms show that higher education and revealed and intended recreation use augment the positive utility effects of an improvement.

The results from all six models can be used to predict average WTP for the water quality improvements. For example, using the formula in equation (12) and sample mean values for *college*, *trplastyr*, and *trpnextyr* we find the following:

- For model 1 the annual WTP for program II is $233, with a 95% confidence interval of ($176, $298).
- For models 2 and 3 the corresponding figures are $173 ($117, $230) and $229 ($173, $293), respectively.[4]

As expected, recoding all uncertain votes to "no" in model 2 leads to a lower mean WTP. Using the formula in equation (15) and the sample means for the interaction variables we find the following:

- For model 4 with $\Delta qual$=0.6, which is equivalent to program II, our mean WTP estimate is $241 per year, with a 95% confidence interval of ($210, $283).
- For models 5 and 6 the corresponding estimates are $195 ($168, $226) and $252 ($220, $296), respectively.

In addition to the results reported in Table 14, we estimated a number of other models. These models indicate that there was no difference in the responses to the two versions of the water quality descriptions (comparing the three-attribute and five-attribute versions). We also ran a Heckman sample selection model (Heckman 1979) using demographic data on Knowledge Networks panel members who were invited to take the survey and declined. Using this approach

[4] Confidence intervals were estimated using the Krinsky and Robb (1986) procedure.

we did not find a statistically significant sample selection bias.

Policy Contingent Valuation Model

As illustrated above the estimates in Table 14 allow us to compute the mean value (WTP) of a specified change in the water quality index using any of the six models. It is important to note that these values should not be compared with similar calculations from the choice experiment. There, the marginal value was for a single trip to a single lake, and the relevant time frame was a single-trip choice occasion. This estimate is for all the lakes in a state, and it is an annual value for the change. Viewed in this light, the two models arguably are consistent with the same underlying preferences and water quality values.

The CV model that we suggest for policy purposes is model 5 in table 14. This model uses the continuous index to measure changes in water quality. By recoding all uncertain responses as “no” votes, it also provides more conservative estimates of WTP than the other models. This model suggests that the benefits of statewide improvements accrue to all residents (perhaps reflecting some types of nonuse value), but that actual and potential recreation uses value the improvements more. This is an intuitive finding, and further supports the validity of our application.

6. Case Study

The objective of this project is to provide a protocol that can be used by state water quality managers to measure the dollar-denominated benefits of proposed numeric nutrient criteria. As described in the introduction, this task requires two types of models: one that can map measures of water quality obtained from a monitoring station network (e.g., total nitrogen, chlorophyll a) to a descriptive quality-level indicator, and one that can map changes in the descriptive quality indicator to dollar values. With the methods described in the water quality modeling technical document (Phaneuf et al., 2009) and the previous sections of this document, the tools that we need to define a protocol are now in place. In this section we demonstrate how these tools can be used by applying them to a case study. For discussion purposes, we define the main policy problem as lake specific. That is, a manager is charged with evaluating a numeric nutrient criterion for a particular lake, and she must assess the recreation benefits of changes from the status quo to the new criterion.

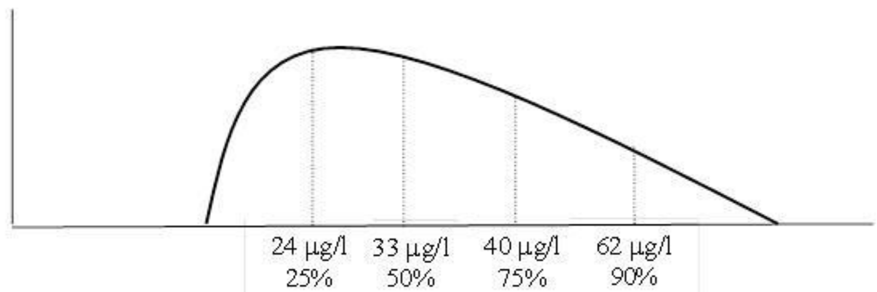
We use an application to valuing nutrient reductions in North Carolina's Falls Lake as our case study example. Nutrient targets for Falls Lake are currently under debate by the North Carolina Division of Water Quality; as a result, a large amount of monitoring station data are available. In particular, median (mean) measures of key nutrient parameters taken throughout 2006 are presented in Table 15. These summaries are based on 270 sampling events. Figure 5 describes in more detail the specific distribution of baseline readings for chlorophyll a.

North Carolina has discussed setting a nutrient criterion such that no more than 10% of chlorophyll a readings are over 40 μg/l. Figure 5 shows that under baseline conditions 10% of readings are over 62 μg/l. Thus, the policy objective that we evaluate in this case study is one that shifts the distribution of chlorophyll a so that the criterion is met (i.e., so that the 90th

Table 15. Median and (Mean) Values for Key Nutrient Parameters

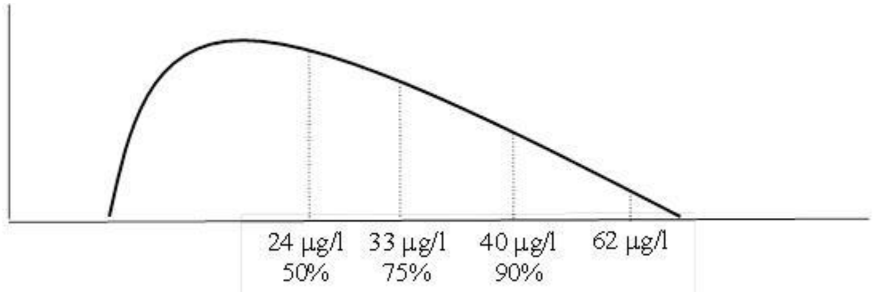
	2006 Baseline Medians and (Means)	Counterfactual Medians
Total nitrogen (TN)	0.76 (0.79) mg/l	0.70 mg/l
Total phosphorus (TP)	0.05 (0.07) mg/l	0.048 mg/l
Chlorophyll a (CLA)	33.00 (35.80) μg/l	24.00 μg/l
Secchi depth (S)	0.70 (0.70) m	0.74 m
Turbidity (T)	9.65 (14.06) NTU	9.05 NTU

Figure 5. Distribution of Chlorophyll a Readings from 2006



percentile of the chlorophyll a distribution is at most 40 μg/l). For this purpose, Figure 6 shows the assumed counterfactual distribution of chlorophyll a, which we have constructed by assuming a leftward shift in the percentile values. Based on this assumption for the counterfactual distribution, we consider the benefits of moving from the 2006 baseline median for chlorophyll a of 33 μg/l to a counterfactual chlorophyll a level of 24 μg/l. With the counterfactual standard for chlorophyll a set, it only remains to impute the implied values of the other nutrients in the counterfactual scenario. This is done using auxiliary regressions estimated using the monitoring network data that separately fit TN, TP, S, and T as a function of CHL. Table 15 presents the full set of nutrient measures for the counterfactual, based on the target for chlorophyll a and the imputed values for the other parameters. Note that these are fairly substantial changes in ambient water quality. In what follows we examine different benefit assessments arising from a change in the baseline to the counterfactual medians.

Figure 6. Distribution of Chlorophyll a Readings Under Counterfactual Assumptions



The first step is to use the tools developed in Phaneuf et al. (2009) to predict the water quality index that corresponds to the nutrient parameter medians. As described in Phaneuf et al. (2009), the actual water quality level in a lake will be influenced by many factors and will vary over time. Lakes with the same values for the nutrient measures listed in Table 16 could end up in different eutrophication categories. Our model first predicts the probability that a lake with a given set of nutrient measurements would be in a given eutrophication category (see Phaneuf et al., 2009 for details, and Kenney, 2007 for background). Using our preferred specification, Table 16 presents the predicted probabilities for each eutrophication category for our baseline and counterfactual scenarios. Based on the results in Table 16, we predict that the expected index level at baseline conditions is 3.45, suggesting that current water quality conditions in Falls Lake lie somewhere between levels C and D. The same prediction using the counterfactual nutrient concentrations results in an expected index level of 3.15. That is, the hypothetical policy intervention improves water quality so that Falls Lake approximately reaches level C.

The second step is to use the choice experiment model estimates to compute the per-trip WTP for the counterfactual improvement. We use equation (9) and the parameter estimates from the Table 12 "all" model so that

$$E(WTP) = \frac{1}{0.016}\begin{bmatrix}(p_A^c - p_A)\times 2.56 + (p_B^c - p_B)\times 1.93 \\ +(p_C^c - p_C)\times 0.98 + (p_D^c - p_D)\times 0.17 - (p_E^c - p_E)\times 0.11\end{bmatrix}, \qquad (16)$$

Table 16. Predicted Baseline and Counterfactual Probabilities for Each Eutrophication Category

Eutrophication Category	Predicted Probability Under Baseline	Predicted Probability Under Counterfactual
Level 1 (Category A)	0.02	0.03
Level 2 (Category B)	0.06	0.12
Level 3 (Category C)	0.43	0.54
Level 4 (Category D)	0.46	0.30
Level 5 (Category E)	0.04	0.02

where the estimated baseline and counterfactual probabilities are computed from the water quality models and shown in Table 16. Using this formula we find an average per-trip WTP of $15.29.

Finally, to obtain an aggregate estimate of the recreation value of the Falls Lake improvement, we need estimates for the number of trips to the lake each year, a policy timeline, and an assumption for the discount rate. Based on the available data on visitation to Falls Lake (see NC DWQ, 2010), we conservatively assume there are 0.9 million trips to the lake each year and the annual benefits from the policy are $13.76 million. If we evaluate the benefits over 20 years and use a 5% discount rate, the present value of the stream of benefits is $171.52 million.

These values are based on several implicit assumptions. For example, the calculation assumes that the new quality level is reached immediately, when in fact it is likely to take several years before ambient conditions in the lake respond to current policy actions. To illustrate the importance of this complexity, consider a 20-year horizon in which water quality stays at baseline conditions for the first 5 years before obtaining the counterfactual level in the sixth year. In this case there are no program benefits in the first 5 years (because water quality has not improved). The annual benefits of $13.76 million per year accrue beginning in the sixth year, implying that the present value of benefits is now $108.81 million for the 20 years of the program.

As part of this project, we developed a spreadsheet tool that allows users to experiment with the protocol used to produce the estimates from this case study. This is available by navigating to http://www.epa.gov/nandppolicy/links.html, and clicking on the ‘grants’ folder once reaching this page. Appendix E contains the user manual that accompanies the spreadsheet tool and explains the equations and data behind the tool (see *User’s Manual for the Water Quality Spreadsheet* in Appendix E); the user manual is also posted at the web site.

7. Conclusion

The intent of our project has been to present a protocol for valuing the nonmarket benefits of numeric water quality criteria. For this purpose we have developed two sets of models: a water quality production function that maps changes in nutrient concentrations to changes in narrated water quality conditions and an economic model that maps changes in narrated conditions to dollar-denominated benefits. Methods related to the former are described in our earlier technical document (Phaneuf et al., 2009). In this document we have focused on describing the data collection and analysis used for the economic modeling.

In general, we find that our models based on the choice experiment and CV data produce intuitive and stable estimates of water quality benefits. Other summaries of behavior, attitudes, and beliefs among survey respondents provide further evidence that our descriptions of water quality in the survey were effective. From this we conclude that our estimates can be applied as one input into the process of evaluating proposed state-level numeric nutrient criteria. For this purpose we have also prepared a spreadsheet that integrates both aspects of our protocol (expert elicitation and the choice experiment analysis) into a non-technical tool that analysts can use. Appendix E to this report contains the user manual for the spreadsheet. The spreadsheet itself and an electronic version of the manual are available at www.epa.gov/nandppolicy/links.html (click on the 'grants' folder once reaching the page).

To close, we offer some caveats and limitations to the benefits estimates provided here. First, our choice experiment models by construction focus on a single aspect of the many ways that water quality improvements can provide economic benefits. Thus, the choice experiment estimates arising from our protocol are likely to be lower bounds in the sense that they do not include nonrecreation benefits of quality changes. Second, we have focused on keeping our economic models simple and transparent in order to make our protocol accessible to nonexperts. Thus, we do not include an endogenous trip-response margin or the effect on visits to substitute sites. In a technical sense this means our protocol computes the benefits of existing trips conditional on the existing allocation of trips to the set of available lakes in a region. If an ad hoc assumption about trip increases due to increases in water quality is not used, this too implies our estimates are likely to be lower bounds. Finally, in our case study we have focused on using the choice experiment model because it provides the most direct link to a well-defined policy objective and potential outcome. The improvement scenario used in the CV model makes the CV estimates most appropriate for a general assessment of statewide water quality standards, or an analysis of a large scale, regional policy intervention.

Limitations

To these caveats, we add a more formal listing of the specific limitations of this study, and those more generally associated with the tools we have applied here. First, in our case study we have relied on a trip choice conjoint experiment to measure the effects of water quality on preferences for lakes. As noted above this limits our focus to use value associated with nutrient pollution reductions. To the extent that there is nonuse value associated with these improvements, estimates from the choice experiment application may be a lower bound on total value. Policies designed to reduce nutrient pollution may also enhance water quality in other dimensions, such as bacteria levels associated with health risks. The economic value of this type of co-benefit is not explicitly included in our estimates. This again suggests our estimates represent lower bounds. Finally, our choice scenarios focus on day trip behavior, since overnight and longer trips tend to be different behavioral phenomena. We advise computing aggregate benefit estimates from this model using total day trip counts, meaning that the benefits to longer stay visitors of quality improvements may not be captured in the aggregate WTP predictions. The sum of these limitations arising from the structure of our preferred behavioral model suggests that the resultant WTP estimates are lower bounds on the value of quality improvements.

Second, we have relied heavily on stated preference techniques to estimate our benefit functions. Stated preference methods are somewhat controversial outside of the environmental economics community given concerns about hypothetical bias – i.e. the potential for respondents to answer questions differently than they would if they were facing a real choice. Our choice experiment estimates are at less risk of hypothetical bias than other types of SP questions, since we focus on a subsample of respondents who have actual experience with the type of choices we are presenting. We followed a careful and thorough approach to designing the study to minimize respondent confusion and we took steps, such as the inclusion of a "cheap talk" script, to reduce the potential for hypothetical bias. The risk of invalidity from hypothetical bias is therefore likely small, though (informal evidence aside) we are not able to provide formal tests to confirm its absence such as comparing the choices to real choices. Our CV estimates are comparatively more likely to be subject to hypothetical bias, in that the choice situation is less familiar to most respondents than a destination choice (and the value concept is less concrete than recreation services). As noted in the discussion of the CV results, our results conform to expectations and common assessments of validity that suggests we have a valid representation of preferences. In addition, our analysis of the uncertainty response indicators is designed to limit any overestimates of value due to hypothetical bias. However, we are not able to formally conclude that people would answer similarly in a real payment situation, since our design did not include this type of real payment treatment.

Third, there are potential limitations arising from our survey design choices. We view our description of the nutrient pollutions levels in Table 1 in non-technical language as a strength of the study. Nonetheless, compromises were necessary in a few instances. For the lay audience we needed to present the attribute levels and their correspondence with the quality categories linearly, though the actual ecosystem might display some non-linearity in how attribute levels relate to overall water quality. The natural brown color of many lakes in the southeastern US created challenges for communicating the color gradation between high and low quality lakes; we ultimately decided to associate brownish color with lesser quality levels, though in reality a brownish tint need not signal poor quality in other dimensions. In addition, our focus groups revealed that many people had not personally experienced large alga blooms or offensive odors at places they had visited. While focus group participants did indicate an understanding of the lower ranges of our quality spectrum, the fact that they had not visited destinations with such low quality may mean our estimates of preferences associated with the worst quality levels are less reliable than estimates for the middle of the quality range. Though we do not view any of these potential limitations as serious, we note that it is not possible to predict the direction of bias that they might cause.

Fourth, our experimental design for the contingent valuation study likely resulted in limited statistical power to detect scope effects among the different program quality levels. Use of a common baseline quality level for all treatments, and limiting the magnitude of quality improvements to physically feasible levels, meant we could not generate wide variability in the quality improvements offered by the different programs. We were therefore not able to estimate statistically significant scope affects with our most flexible model, though with some additional structure the evidence for scope effects is robust.

Fifth, our modeling choices necessarily imply some limitations in that the quantitative predictions depend on these choices (this is less so for qualitative predictions). For example, we have used a log specification in the contingent valuation model to impose intuitive, but non-testable, structure on the preference function. More importantly, our construction of the price of a visit to a lake in the choice experiment includes assumptions on the opportunity cost of time, travel speed, and out of pocket travel costs. Our choices for these parameters are consistent with what is used in the recreation demand literature, and are generally conservative (i.e. our assumptions imply lower travel costs than other studies that use, for example, the full wage rate as the opportunity cost of time). Our estimates of value are dependent on these choices and they would be different if we made alternative decisions. Given that our choices were conservative any errors arising from them likely imply predictions from the model are an under estimate of value – a direction consistent with the first limitation described above.

Sixth, the KN sample design is nationally representative, but this does not carry over to individual states. Thus, our sample of KN respondents from, say, North Carolina may not be representative of the population in that state. Our comparison of census and state sample averages for common household characteristics confirms that there are some differences between the target and sample populations. Though this is a point of concern, it is worth noting that our estimates suggest that household characteristics such as income, family size and status, and age are not strong predictors of the behavior we are studying.

Our final limitation concerns the transferability of our estimates. The protocol we have designed is intended for use by state water quality managers in the southeastern US. We note, however, that the water quality model was calibrated using North Carolina specific data. Likewise, the survey sample focused most heavily on the core study area of North Carolina, South Carolina, and Virginia. The accuracy of predictions from our protocol is likely to decrease with distance from the core study area, and the validity of our water quality modeling will be diminished when the policy lakes differ in hydrology from the North Carolina environment.

Closing Remark

To our list of caveats and limitations, we close by adding the obvious comment that any model is only as good as the data that are fed into it. Analyses of the type presented in our case study depend on having quality data on baseline conditions at a policy lake, a good prediction of the counterfactual scenario, information on aggregate visits to the policy lake, and a willingness to engage in sensitivity analysis to understand the uncertainties in the predictions. When properly deployed and interpreted in this way, our sense is that the protocol we have developed can serve as a valuable input into state-level water quality policy evaluation.

8. References

Banzhaf, H.S., D. Burtraw, D. Evans, and A. Krupnick. 2006. "Valuation of Natural Resource Improvements in the Adirondacks." *Land Economics* 82(3):445-464.

Blumenschein K.**,** Blomquist G.C., Johannesson M., Horn N., Freeman P. 2008. Eliciting Willingness to Pay without Bias: Evidence from a Field Experiment. *The Economic Journal*; 118(January): 114-137.

Boyle, K.J. 2003. "Contingent Valuation in Practice." In *A Primer on Nonmarket Valuation,* P.A. Champ, K.J. Boyle, and T.C. Brown, eds. Dordrecht: Kluwer Academic Press.

Carson, R.T., and T. Groves. 2007. "Incentive and Informational Properties of Preference Questions." *Environmental and Resource Economics* 37:181-210.

Cesario, F.J. 1976. "Value of Time in Recreation Benefit Studies." *Land Economics* 52:32-41.

Champ, P.A., R.C. Bishop, T.C. Brown, and D.W. McCollum. 1997. "Using Donation Mechanisms to Value Nonuse Benefits from Public Goods." *Journal of Environmental Economics and Management* 33(2):151–62.

Farber, S., and B. Griner. 2000. "Valuing Watershed Quality Improvements using Conjoint Analysis." *Ecological Economics* 34:63-76.

Freeman, A.M. 2003. *The Measurement of Environmental and Resource Values*," 2nd edition. Washington, DC: Resource for the Future Press.

Greene, W. 2000. *Econometric Analysis, 4th edition*, Upper Saddle River, NJ: Prentice-Hall.

Heckman, J. 1979. Sample selection bias as a specification error. *Econometrica* 47: 153--161.

Hensher, D.A., J.M. Rose, and W.H. Greene, eds. 2006. *Applied Choice Analysis: A Primer*. Cambridge: Cambridge University Press.

Herriges, J.A., C.L. Kling, and J. Tobias. 2010. “What are the Consequences of Consequentiality?” *Journal of Environmental Economics and Management* 59:67-81.

Holmes, T.P., and W.L. Adamowicz. 2003. “Attribute-Based Methods.” In *A Primer on Nonmarket Valuation*, P.A. Champ, K.J. Boyle, and T.C. Brown, eds. Dordrecht: Kluwer Academic Press.

Kenney, M.A. 2007. “Which Nutrient Criteria Should States and Tribes Choose to Determine Waterbody Impairment? Using Science and Judgments to Inform Decision Making.” Duke University Dissertation.

Kling, C., D. Phaneuf, and J. Zhao. 2008. “From Exxon to BP: has some number become better than no number?” *Journal of Economic Perspectives* 26(4): 3-26.

Krinsky, I. and A.L. Robb. 1986. "On Approximating the Statistical Properties of Elasticities." *Review of Economic and Statistics* 68: 715-719.

Jeanty, P. W. 2007. "wtpcikr: Constructing Krinsky and Robb Confidence Interval for Mean and Median Willingness to Pay (WTP) Using Stata." *North American Stata Users' Group Meetings 2007*, 8.

Leggett, C.G., and N.E. Bockstael. 2000. “Evidence of the Effects of Water Quality on Residential Land Prices.” *Journal of Environmental Economics and Management* 39:121-144.

Mathews, K.E., M.L. Freeman, and W.H. Desvousges. 2007. “How and How Much? The Role of Information in Stated Choice Questionnaires.” In *Valuing Environmental Amenities Using Stated Choice Studies*, B.J. Kanninen, ed. Dordrecht: Springer.

National Oceanic and Atmospheric Administration (NOAA). 1993. “Natural Resource Damage Assessments under the Oil Pollution Act of 1990.” *Federal Register* 58:4601-4614.

North Carolina Division of Water Quality (NC DWQ). 2010. *Fisk Analysis for Proposed Nutrient Strategy for Falls of Neuse Reservoir*.

Olmstead, S.M. 2010. "The Economics of Water Quality." *Review of Environmental Economics and Policy* 4:44-62.

Parsons, G.R. 2003. "The Travel Cost Model." In *A Primer on Nonmarket Valuation,* P.A. Champ, K.J. Boyle, and T.C. Brown, eds. Dordrecht: Kluwer Academic Press.

Phaneuf, D.J. 2002. "A Random Utility Model for TMDLs: Estimating the Benefits of Watershed Based Ambient Water Quality Improvements." *Water Resources Research* 38.

Phaneuf, D.J., and V.K. Smith. 2005. "Recreation Demand Models." In *Handbook of Environmental Economics*, K. Maler and J. Vincent, eds. Amsterdam: North Holland.

Phaneuf, D.J, M. Kenney, and K. Reckhow. 2009. *Measuring Nutrient Reduction Benefits for Policy Analysis Using Linked Non-Market Valuation and Environmental Assessment Models—An Interim Report on Water Quality Modeling.* Report submitted to U.S. EPA Office of Water.

U.S. Environmental Protection Agency (EPA), Office of Water. 2007. "FY2007 Nutrient Benefits Valuation." Request for proposals.

Viscusi, W.K., J. Huber, and J. Bell. 2008. "The Economic Value of Water Quality." *Environmental and Resource Economics* 41:169-187.