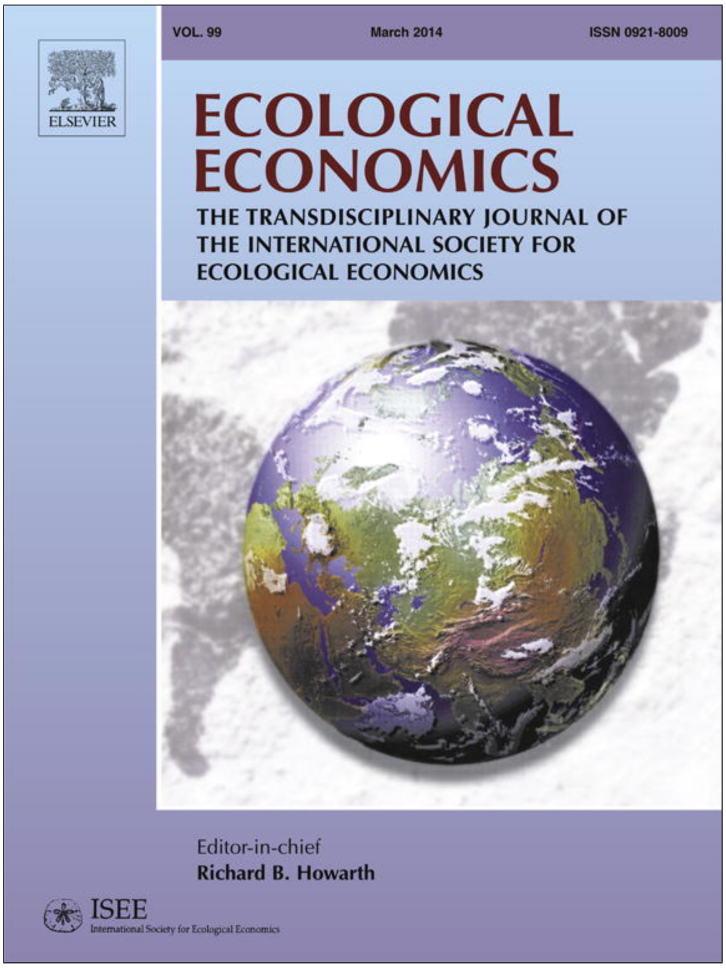
VOL. 99
March 2014
ISSN 0921-8009
ELSEVIER
ECOLOGICAL ECONOMICS
THE TRANSDISCIPLINARY JOURNAL OF THE INTERNATIONAL SOCIETY FOR ECOLOGICAL ECONOMICS
Editor-in-chief
Richard B. Howarth
ISEE
International Society for Ecological Economics

Ecological Economics 99 (2014) 40–52

Contents lists available at ScienceDirect

Ecological Economics

journal homepage: www.elsevier.com/locate/ecolecon

Analysis

Combining expert elicitation and stated preference methods to value ecosystem services from improved lake water quality

George Van Houtven [a], Carol Mansfield [a], Daniel J. Phaneuf [b,*], Roger von Haefen [c], Bryan Milstead [d], Melissa A. Kenney [e], Kenneth H. Reckhow [f]

[a] RTI International, United States
[b] University of Wisconsin, United States
[c] North Carolina State University, United States
[d] U.S. Environmental Protection Agency, United States
[e] University of Maryland, College Park, United States
[f] Duke University, United States

ARTICLE INFO

Article history:
Received 18 March 2013
Received in revised form 23 December 2013
Accepted 31 December 2013
Available online 31 January 2014

Keywords:
Ecosystem services valuation
Water quality
Expert elicitation
Stated preference
Water quality modeling

ABSTRACT

With increasing attention on the contribution of ecosystems to human well-being, there is a need for tools that integrate ecological and economic models for valuing ecosystem services. To address this, we develop a protocol for linking ecological processes and outcomes to human preferences, which combines environmental modeling, expert elicitation, and nonmarket valuation methods. Our application values reductions in nutrient loads to lakes in the southeastern US. The innovation centers on how biochemical measures of water quality (e.g., chlorophyll *a*) are translated into terms that are meaningful to individuals who derive ecosystem services from them. Using expert elicitation data, we estimate a model linking changes in biochemical measures to an index of eutrophication in lakes. We then develop a stated preference survey including (a) detailed descriptions of the perceptible outcomes – e.g., water color, clarity – associated each eutrophication index level; and (b) policy scenarios involving state-level changes in lake eutrophication conditions. We estimate a function that predicts households' willingness to pay for changes in lake water quality. We demonstrate the protocol through a case study examining the benefits of lake quality improvement in Virginia as a result of recent policies to reduce nutrient loads in the Chesapeake Bay watershed.

1. Introduction

Ecosystems provide valuable services to households and businesses, but a number of challenges make it difficult to estimate the monetary value of these services. While economists have developed an impressive collection of methods for valuing nonmarket and public goods, practical applications that seek to value changes in ecosystem services face additional challenges. For example, quantitative measurements of ecosystem health (such as nutrient concentrations in surface water) are typically not good descriptors of the actual services that people perceive and derive value from. Nonetheless regulators usually set management goals based on chemical, physical, or biological properties of the resource. This creates disconnects between how ecosystem quality is assessed, how ecosystem services are defined, and the way that economists go about measuring the value of these services. Many otherwise carefully executed ecosystem service valuation studies do not deliver on their policy promise, owing to this difficulty in precisely linking changes in the valued services to the physical outcomes of a policy shock.

Our research addresses this problem by developing an integrated ecosystem services valuation protocol that connects changes in ecosystem health indicators to changes in economic value in a way that maintains direct linkages between physical measures, service levels, and household preferences. Our specific application values reductions in nutrient loadings to freshwater lakes in the southeastern United States. The US Environmental Protection Agency (EPA) has encouraged states to set numeric criteria for nitrogen, phosphorus, and chlorophyll *a* concentrations as a way of controlling eutrophication (Kenney et al., 2009; Reckhow et al., 2005; USEPA, 2010). Jurisdictions must also develop Total Maximum Daily Load (TMDL) limits for impaired waters. The economic benefits that these ambient standards and TMDLs provide, however, arise from people's preferences and the underlying services they receive. Thus while quantitative indicators are invaluable for assessing ecosystem health and establishing policy objectives, the benefits they provide can be difficult to conceptualize. In contrast descriptive narratives of quality improvements are useful for communicating the possibility of benefits, but their imprecise nature is what led the EPA to encourage the development of numeric criteria in the first place.

In this paper we present an approach that combines water quality modeling, expert elicitation, and a stated preference survey to quantify the linkages between changes in nutrient loadings, changes in ambient

* Corresponding author at: Department of Agricultural and Applied Economics, University of Wisconsin-Madison, United States.
E-mail address: dphaneuf@wisc.edu (D.J. Phaneuf).

0921-8009/$ – see front matter
http://dx.doi.org/10.1016/j.ecolecon.2013.12.018

concentrations, changes in ecosystem service levels, and ultimately nonmarket values for quality improvements. Expert elicitation is a formal process used to systematically elicit and quantify judgments from scientific experts. In our case we used expert elicitation to collect data from water quality experts and estimate a functional relationship linking observations of specific ambient water quality parameters in the study area lakes to a five-level, multi-attribute eutrophication index. Because eutrophication manifests itself differently in different systems, we used expert elicitation to specifically establish this relationship for reservoirs in the Southeastern region of the United States. Each level of the index was described to experts using narrative descriptions for a defined set of attributes, including clarity, color, algae, aquatic life, and odor. Then, to understand how the public values different eutrophication states and quantify their willingness to pay for less-eutrophic conditions, we applied a stated preference approach. Our stated preference survey took the narrative descriptions from the expert elicitation and modified them to be more accessible to a non-technical audience and, at the same time, consistent with the descriptions presented to experts. Although numeric water quality indexes have been widely used in previous economic valuation studies (Van Houtven et al., 2007), one of the main advantages of this index approach is the direct connection it provides between experts' and lay persons' understanding of water quality. Importantly, the year-long development of our stated preference survey involved substantial back and forth between the water quality experts who developed and conducted the expert elicitation and reviewed the stated preference scenarios, the economists who were translating their knowledge into survey-appropriate descriptions, and individuals from the lay public who were involved in focus groups and pretest interviews. This combination of expert elicitation and stated preference surveying offers an innovative approach for linking changes in chemical and biological water quality parameters, which are typically used as measurable indicators of ecosystem health, with attributes that are more closely linked to the types of ecosystem services that individuals recognize and value in lake water quality. It also provides an approach for linking water quality and preferences that we expect will be more explicit and transparent for policy makers.

The remainder of the paper is organized as follows. In Section 2 we place our research in context by reviewing background information that is relevant for ecosystem services valuation as related to water quality. In Section 3 we present our analytical framework in three subsections that describe (a) the water quality models; (b) expert elicitation analysis; and (c) the process used to translate the experts' understanding into the descriptions used in the survey. It also presents the details of our survey and econometric models, and Section 4 contains a case study. The policy context for our case study is the recently promulgated TMDL limits established by EPA for the Chesapeake Bay, which has received considerable regulatory and media attention. In addition to improving conditions in the Bay estuary, the rule is expected to reduce nutrient loads and improve water quality throughout the Bay's 64,000 square mile watershed. We examine the benefits of the expected lake water quality improvements in the state of Virginia, much of which lies within the Bay watershed. We find that the Chesapeake TMDL will improve lake water quality in Virginia by an amount sufficient to generate $184 million per year in aggregate benefits for residents of the state. Although this estimate is of policy interest, the main objective of the case study is to illustrate a common ecosystem service valuation problem, and to demonstrate the advantages of our approach for addressing it. The paper's main contribution therefore is the development of an integrated protocol combining expert elicitation and stated preference techniques, which would be of use in many practical valuation contexts. We conclude the paper in Section 5 by discussing in greater detail the potential of our approach to advance the practice of ecosystem service valuation generally.

2. Background

The basic ecosystem service valuation problem we address is illustrated by Fig. 1, which traces how a change in an environmental input filters through the system to produce a change in human well-being. Note that the change in actual services and behavior (box 3) is preceded by physical changes that are not generally observed by households. For example, the process begins with a shock to an environmental input to the ecosystem, such as nitrogen loading in our study (box 1). This produces a physical change in the ecosystem (box 2), which is measured by an indicator such as nutrient concentrations in the water. Scientific assessment and regulatory decisions are usually based on the information in box 2, but this is still a secondary outcome for purposes of environmental valuation. It is the perceptible change in the ecosystem and the resulting change in the quantity or quality of services derived from the ecosystem (shown in box 3) that directly impact human well-being. In the case of nutrients, the perceptible ecosystem changes relate to observable features of water bodies, such as color, clarity, smell, and abundance of aquatic life. Box 4 illustrates the final step linking a change in services to preferences and monetary value.

Many studies have addressed a subset of the individual steps shown in Fig. 1. However, relatively few have developed protocols that formally link all four components. For example, there is a large literature applying stated preference methods to value changes in water quality (see Johnston et al., 2005; Van Houtven et al., 2007 for summaries of this literature). Because the water quality changes described in the surveys must be expressed in terms that are understandable to a non-technical audience, they are often non-specific in their correspondence to measurable biophysical parameters. A good example of this is the lake visitation choice experiment used by Roberts et al. (2008), which includes an attribute for the presence/absence (and risk) of an algae bloom at the destination. The study addresses boxes 3 and 4 quite effectively, but by abstracting from boxes 1 and 2 it does not allow policy analysis of how changes in nutrient levels map to changes in the likelihood that a bloom will appear. Other studies (e.g. Egan et al., 2009) have used revealed preference methods to directly link measured water quality to behavior. This approach connects boxes 2 and 4, thereby leaving latent the process by which individuals translate ecosystem quality into ecosystem services. While this strategy is attractive in its ability to directly connect policy targets to valuation, identification and interpretation challenges can be substantial due to uncertainty about the connections underlying the reduced form relationship. Finally, several studies have employed an approach in which multiple pollution parameters are aggregated to a one dimensional index of water quality (USEPA, 2002, 2009a,b). The best known technique characterizes quality along the 0 to 100 scale, based on the results of an expert elicitation

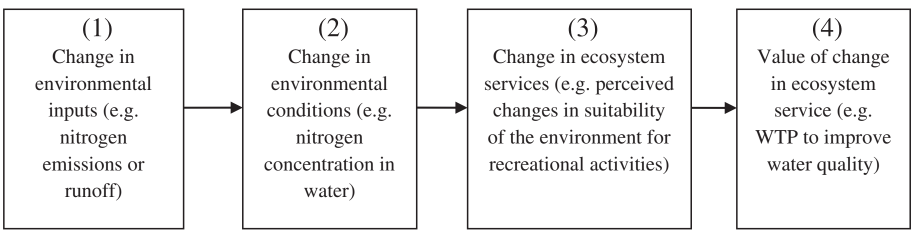

Fig. 1. Measuring the value of a change in ecosystem services.

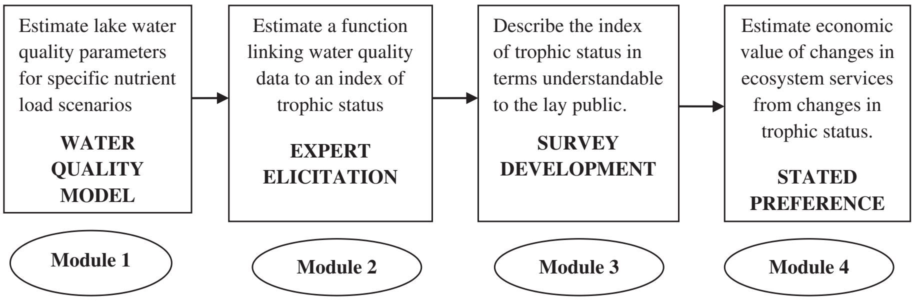

Fig. 2. Overview of approach.

study by McClelland (1974). To apply this 100-point index for water quality valuation, Mitchell and Carson (1984) developed a water quality ladder approach (see, for example, Smith and Desvousges (1986) and Edwards and Anderson (1987) for other applications of this ladder). This approach specifies the minimum index levels at which water quality is suitable for specific human uses — i.e., boatable, fishable, and swimmable. However, one of the main limitations of this approach is that the interpretation of these primarily recreation-based use categories can be fairly subjective, in that different people may associate different water quality conditions (or even a different ranking of conditions) with these activities. Instead of defining biophysical attributes that are understandable to the lay public and then allowing individuals to decide how these attributes affect the ecosystem services they care about, the approach directly defines the uses and implied ecosystem services. Thus the link between boxes 2 and 3 may be non-unique (i.e., individuals may have different interpretations of the water quality changes being valued).

For this study we developed a water quality indexing system for nutrient-related pollution in lakes that describes five categories of eutrophication.[1] A key advantage of this indexing system over previous applications was that it was specifically designed so that it could be linked backward (using the results of an expert elicitation) to a set of commonly monitored and modeled water quality parameters and forward (using a stated preference survey) to human values through verbal and visual descriptions of biophysical water quality conditions. Importantly, these conditions were carefully matched to the descriptions seen by experts in the elicitation exercise. The index was therefore used to connect boxes 2 and 3 in Fig. 1, by linking outputs from the expert elicitation to the descriptions used in the stated preference survey.

In our stated preference survey, we then scaled this index system to the state level by examining the percentage of lakes in the state that fall into each category, and we analyze the stated preference data by valuing changes in the expected index value, which we computed using survey-induced variation in the distribution of lakes in each eutrophication category. While this approach draws on aspects of other stated preference studies that have focused on changes in lake water quality in the US (Banzhaf et al., 2006; Herriges et al., 2010; Viscusi et al., 2008), it is to our knowledge the only study to integrate and operationalize all the steps in Fig. 1 in a way specifically designed to maintain consistency among the various elements.

3. Analytical Framework

Our approach to valuing changes in ecosystem services combines water quality modeling and expert elicitation with stated preference survey methods. Fig. 2 illustrates our main objective and provides an overview of the steps we use. Specifically, we want to map different nutrient load scenarios to indicators of surface water quality, and then link these indicators to ecosystem services and finally economic values. To accomplish these objectives, we applied the four main modules shown in Fig. 2, which generally correspond with the four elements (boxes) shown in Fig. 1. These modules are described in this and the next section.

3.1. Module 1 — Water Quality Model

A number of existing water quality models can be used to estimate lake water quality conditions resulting from alternative nutrient load scenarios. Rather than developing a new model, the main objective of module 1 was to select an existing model that is compatible with the other aspects of our protocol. To map into our expert elicitation the main criterion was that the water quality model needed to estimate concentrations of total nitrogen, total phosphorus, and chlorophyll *a* as a function of basin-wide nutrient loads. For compatibility with our main valuation task the model needed to operate at a scale allowing prediction for most of the main lakes and reservoirs within a given state. Finally, to demonstrate the utility of our approach for policy purposes we needed a model that could incorporate nutrient load estimates based on specific policy scenarios.

Based on these considerations, we apply the SPARROW (Moore et al., 2011) model, which uses non-linear regression to develop an empirical relationship between long term mean nutrient flux (kg/yr), predictor variables such as estimated nutrient inputs (based on land cover from the 2002 National Land Cover Database and location), soil type, precipitation, and estimated loss factors (land to water delivery fractions and in-stream/in-basin attenuation). Additional details on our specific use of the SPARROW model for our Chesapeake Bay watershed application are included in Section 5.

3.2. Module 2 — Expert Elicitation

The goal of this module is to estimate a function linking data for multiple water quality parameters in freshwater lakes to an ordinal index of eutrophic conditions. Specifically, the function will take as input a vector of water quality parameters obtained from a monitoring network or modeling exercise, and produce as output a predicted value for the eutrophication index. Characteristics of eutrophication such as increased algal growth, reduced water clarity, coloration of surface water, unpleasant odors, and impacts on aquatic life are perceptible to users and therefore influence decisions and the economic value derived from a lake. However, no single variable is always the best predictor of eutrophication or trophic status because trophic state manifests itself differently in different locations. Because there are not established models for man-made lakes in the Southeastern region, expert elicitation helps to assure that we use scientific expert judgment to map multiple water quality variables to these eutrophication categories in the region of interest.

To develop the function we use data from an expert elicitation conducted by Kenney (2007). Kenney conducted the elicitation with 14 experts familiar with North Carolina lake water quality. These experts were asked to consider how values for seven water quality parameters presented in a controlled experiment map into five levels of trophic

[1] Eutrophication is a process that occurs in nutrient enriched lakes, which results in algal growth, reduced water clarity, coloration of surface water, unpleasant odors, and impacts on aquatic life.

Table 1
Trophic State/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

status. Specifically, the experts were asked to consider how concentrations of total nitrogen, total inorganic nitrogen, total phosphorus, chlorophyll *a*, and dissolved oxygen, as well as turbidity and Secchi depth, map into the five pre-defined trophic state levels described in Table 1.

Each expert was presented with rows of water quality data for the seven parameters, and for each row the expert was asked the following question[2]:

> Imagine 100 different lakes in the (named NC) eco-region with the characteristics specified by the given data row. Of the 100 lakes, how many of the lakes would you expect to fall into each of the (five categories of eutrophication)?

Through this question experts were asked to draw on their professional understanding of eutrophication processes and make their best scientific judgments regarding the specific connection between (a) water quality parameters; and (b) eutrophic water quality outcomes as described by the attributes in Table 1. Thus the task was narrowly focused on the participants' area of expertise. The experts were each given 100 rows of data that were designed to reflect realistic combinations of parameter values, most of which were actual observations taken from reservoirs in the state. Each expert responded to 50 data rows that were the same across all experts and 50 data rows that varied across experts according to the specific eco-region with which they were most familiar.

The actual elicitation process unfolded in three steps, in accordance with accepted best practice (Morgan and Henrion, 1990). Each step was presented separately and all parts were conducted for all the experts who participated. The first part included a semi-structured interview about eutrophication processes and designated use impairment, as well as a discussion on the use of expert judgment in the project. In the second part, the expert provided his or her judgments on the data rows described above. The elicitor worked through the first few cases with the expert until the elicitor judged that the expert understood the elicitation survey, could answer the questions in a manner consistent with their expert understanding, and felt comfortable working through the remaining cases individually. The third step involved follow up and debriefing. For example, experts were asked to look at a subset of their data rows and describe why they made a particular assessment. If the expert saw an error in their assessment, he or she was encouraged to make a correction to more accurately reflect her belief about the trophic state category. The three steps in total took 6 to 8 h of the expert's time. An example of how a single expert responded to a single row of data is shown in Table 2.

The expert elicitation provided 1400 rows of explanatory and response variables similar to the example in Table 2. The experts' rankings are an example of ordinal data — discrete outcomes that have a natural ranking but for which there is no meaningful scale. To summarize expert responses across the 1400 observations, we first select the rank that received the highest proportion of the expert's distribution. Across the full sample of responses the average rank is approximately three, and a standard error of over one suggests there is usable variability in the data among experts' responses. At the same time, there is evidence of consistency in responses across experts. For example, for the 700 observations where all of the experts saw the same 50 rows of water quality data, the average pairwise correlation between the 14 experts was 0.57 and over 72% of these correlations were above 0.5.

The objective of this module is to use statistical modeling to relate the values of the water quality parameters to the full ordinal outcomes provided by experts, and to estimate a functional relationship that can be used as a predictive model for the ordinal outcomes. To address this objective, we apply an ordered logit model, which is a commonly used approach for analyzing ordinal data and has been used in other expert elicitation analyses related to water resources (e.g., Sonneveld and Albersen, 1999). We also explored other modeling approaches such as structural equation modeling and binomial regressions (Kenney, 2007; Phaneuf et al., 2009); however, we concluded that the ordered logit method was best suited for meeting the objectives of this study.

To estimate our ordered logit models, a transformation of the raw data is needed. Recall that our experts provided a *distribution* (i.e. the proportion of lakes that would fall into each category) of ordinal ranks, rather than a single rank, as is needed for this model. We used the full distribution of experts' responses by expanding each response into observations equal to the number of categories with positive (i.e., non-zero) distribution mass. Each of these 3918 observations was given an outcome value corresponding to a category with positive mass, and sample weights based on the proportion of lakes assigned to that category.[3]

Table 3 presents results from three specifications of the weighted ordered logit model. In all instances the standard errors shown in parentheses are clustered to reflect the likely correlation among judgments from the same expert. Model 1 contains all seven of the water quality variables presented to the experts. The coefficient estimates are not comparable in magnitude, but their signs do have a direct interpretation. A positive sign suggests that a higher level of the variable pushes the ranking higher – towards a worse trophic status – and a negative coefficient means that higher levels of the variable are associated with a better (lower index number) trophic status. Based on this interpretation all of the estimated coefficients have sensible signs. As regards statistical performance, most coefficient estimates are significant at a 5% level, with the exception of both nitrogen types and dissolved oxygen. The former likely suggests total nitrogen and total inorganic nitrogen were redundant sources of information for the experts. The latter reflects the fact that surface dissolved oxygen (the variety presented) is less relevant than other oxygen measures for predicting trophic status. A Wald test that the coefficients on total inorganic nitrogen and dissolved oxygen are jointly zero fails to reject the null hypothesis ($\chi^2(2) = 3.24$, p-value $= 0.197$).

Given these findings, model 2 examines coefficient estimates when inorganic nitrogen and dissolved oxygen are excluded. We find that all remaining coefficients are now significant, and the Akaike Information Criteria (AIC) statistics suggest only a small loss in model fit.[4] Model 3 is motivated by the practical requirements of linking this model to

[2] Each expert answered this question for the specific North Carolina eco-region in which they were most familiar. The eco-regions include Coastal (4 experts), Southeastern Plains (1 expert), Piedmont (6 experts), and Blue Ridge (3 experts).

[3] For example, for the response shown in Table 2 three observations are derived: one each recording index values 2, 3, and 4. The weights assigned to these in estimation are 0.10, 0.50, and 0.40, respectively. Additional discussion of the expert elicitation data and our different modeling approaches is contained in Phaneuf et al. (2009).

[4] For model 1 $AIC_1 = 9856.22$, and for model 2 $AIC_2 = 9857.89$.

Table 2
Example of expert response to main elicitation task (in italics).

Parameter:	Photic total nitrogen	Photic total inorganic nitrogen	Photic total phosphorus	Photic chlorophyll *a*	Surface dissolved oxygen	Secchi depth	Photic turbidity
Value:	0.46 mg/l	0.02 mg/l	0.03 mg/l	38 μg/l	6.3 mg/l	1.3 m	3.9 NTU

Ranking on trophic index	1	2	3	4	5
Number of lakes	*0*	*10*	*50*	*40*	*0*

SPARROW model outputs, which do not include Secchi depth or turbidity. When these two variables are excluded, we find that all three remaining coefficients are intuitively signed and significant; however, a Wald test suggests a statistically significant loss of explanatory power from their exclusion ($\chi^2(2) = 3.54$, p-value < 0.001). In all models, we can also reject with a high level of significance ($p < 0.01$) the hypotheses that any pair of "cut" threshold values were equal, which indicates no evidence to support combining two or more trophic categories into one.

The results in Table 3 allow us to predict a lake's trophic index based on the values of selected water quality parameters. In particular, model 2 is a function with four inputs (concentration levels of total nitrogen, total phosphorus, chlorophyll *a*, and turbidity) that produces as output the probabilities $p_1,\ldots,p_5$, which describe the likelihood that the lake is in each of the five categories given in Table 1, conditional on the values for the four inputs. For example, when we use the concentrations of nitrogen, phosphorus, and chlorophyll *a* and turbidity shown in Table 2 as input, our model predicts the following probabilities: $p_1 = 0.038$, $p_2 = 0.116$, $p_3 = 0.5044$, $p_4 = 0.322$, $p_5 = 0.021$. According to these results, the lake would have the highest probability of being in the trophic category 3.

3.3. Module 3 — Survey Development

In addition to understanding how experts relate observed water quality data to trophic state categories, we needed to understand the public's preferences for improved water quality. For this we designed and implemented a stated preference survey to estimate households' values for improving nutrient-related water quality in lakes in their home state. We describe the survey in general in the following section. Here we focus on one of its key elements: our translation of the main elements of Table 1 into terms that would be meaningful to the general public, while also maintaining consistency with the experts' scientific understanding of the five trophic status levels. To accomplish this we focused on five attributes of lakes in the southeastern US – color, clarity, fish, algal blooms, and odor – that we concluded were the most relevant indicators of the final ecosystem services conveyed by lakes to individuals. This conclusion was based on three separate focus groups, which confirmed that these attributes are (a) directly observed and easily understood by the general public; and (b) salient to individuals in the sense that their levels influence preference rankings of different lakes. We did not include the attributes 'nutrient levels' or 'oxygen' because we judged them to be largely unobservable and intermediate to the features of lakes over which preferences are defined.

Using focus groups and the simultaneous advice of water quality scientists we developed descriptions of the five main attributes – color, clarity, fish, algal blooms, and odor – and defined discrete levels for each attribute. We then used the attributes and their levels to define five water quality categories, which corresponded directly to the five trophic index levels used by the water quality experts. Table 4 shows the summary level descriptions included in the survey, where the quality levels are labeled A (best) to E (worst) and correspond to the rankings 1 (least trophic) to 5 (most trophic) in Table 1. The five quality levels in Table 4 constitute our definitions of how the quality of services provided by freshwater lakes is impacted by the underlying health of the ecosystem. To promote respondents' understanding of the index, the online survey first included separate descriptions for each of the five attributes and their levels, before grouping them into index categories as in Table 4. Photographs were used to show variations in lake color and clarity, as well as different sizes and locations of algae blooms. In addition, to further encourage respondents to examine and consider the meaning of each quality (trophic) level, the survey asked them to select the quality categories that they believed were most common for lakes in their state.[5]

It is important to note that throughout the stated preference survey, we maintained the direct connection between water quality categories (A through E) and water quality attributes (color, clarity, etc.) as shown in Table 4. In other words, the water quality attributes always varied together in the pattern shown in Table 4 and never independently from one another. This design feature was selected for the following reasons. First, the attributes are highly correlated in practice. To maintain realism any designed variation in the attributes would therefore have been constrained to lie within a limited range. Second, it maintains direct and important consistency with the expert elicitation design and analysis. In particular, Module 2 does not accommodate lake attribute combinations that differ from the patterns shown in Tables 1 and 4. Third, it reduced the cognitive burden on respondents to have them focus on

Table 3
Ordered logit regression analysis of responses to expert elicitation.

	Model 1	Model 2	Model 3
Total nitrogen	0.282	0.389**	1.244***
	(0.199)	(0.169)	(0.195)
Total inorganic nitrogen	0.582*		
	(0.339)		
Total phosphorus	7.401***	7.836***	9.289***
	(1.726)	(1.787)	(1.924)
Chlorophyll *a*	0.0590***	0.0588***	0.0574***
	(0.00703)	(0.00652)	(0.00622)
Dissolved oxygen	0.00723		
	(0.0454)		
Secchi depth	−0.552***	−0.537***	
	(0.0924)	(0.0939)	
Turbidity	0.0152**	0.0171***	
	(0.00675)	(0.00620)	
Cut 1	−0.936	−0.944**	0.241
	(0.598)	(0.436)	(0.314)
Cut 2	0.689	0.681*	1.775***
	(0.472)	(0.348)	(0.286)
Cut 3	2.651***	2.645***	3.654***
	(0.410)	(0.340)	(0.322)
Cut 4	4.940***	4.925***	5.935***
	(0.438)	(0.388)	(0.405)
Log pseudolikelihood	−4917.11	−4919.95	−5035.82
Pseudo R-squared	0.1645	0.1640	0.1443
Observations	3918	3918	3918

Standard errors are in parentheses.
* $p < 0.10$.
** $p < 0.05$.
*** $p < 0.01$.

[5] Phaneuf et al. (2013) contains a detailed description of the survey development process. To view the survey instrument and technical documents associated with the project go to www.epa.gov/nandppolicy/links.html, and click on the 'grants' folder for access to the 'Nutrients Benefits Valuation' project information.

Table 4
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 ft deep or more	Can see 2–5 ft deep	Can see 1–2 ft deep	Can see at most 1 ft deep	Can see at most 1 ft 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.

five types of lakes rather than on a much larger variety of lakes with varying attributes.

3.4. Module 4 — Stated Preference Analysis

The sample frame for our stated preference survey was households in the eight southeastern states that contain the largest portion of EPA's Nutrient Eco-region IX (USEPA, 2000): Alabama, Georgia, Kentucky, Mississippi, North Carolina, South Carolina, Tennessee, and Virginia. Within this region, we consider three states – North Carolina, South Carolina, and Virginia – to be the core of our study area, because it is within this region that we expect module 2 (the expert elicitation) to provide the most reliable predictions. We used the market research firm Knowledge Networks (KN) to execute a web-based survey using a nationally-representative sample from members of their household panel.[6]

The survey vehicle contained several sections. Our focus here is on a contingent valuation (CV) application that was one of two major valuation components that were included. The objective of the CV component was to measure people's willingness to pay (WTP) to shift the distribution of lake water quality in their state from baseline conditions to improved conditions, based on a hypothetical public program.[7] After receiving information on nutrient pollution and the categories of lakes shown in Table 4 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.

Accomplishing the valuation task required that we communicate both the baseline (without the program) conditions and conditions resulting from a policy change (with the program). We used the following text and graphical format to describe the change:

> The diagram below compares projected lake conditions in HOME STATE in 10 years, 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. X% 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.

The same baseline condition was presented to all respondents; however, four different versions of the 'with program' description were randomly assigned and presented to respondents. Table 5 shows the distributions that were presented in each version, and Fig. 3 provides an example of the graphic respondents received.

Table 5
State-level distributions of water quality levels shown in the CV survey.

Trophic category/index	"Without the program"	"With the program" versions			
	Baseline	I	II	III	IV
A/1	5%	10%	15%	10%	20%
B/2	25%	25%	35%	55%	45%
C/3	50%	50%	40%	30%	30%
D/4	15%	15%	10%	5%	5%
E/5	5%	0%	0%	0%	0%

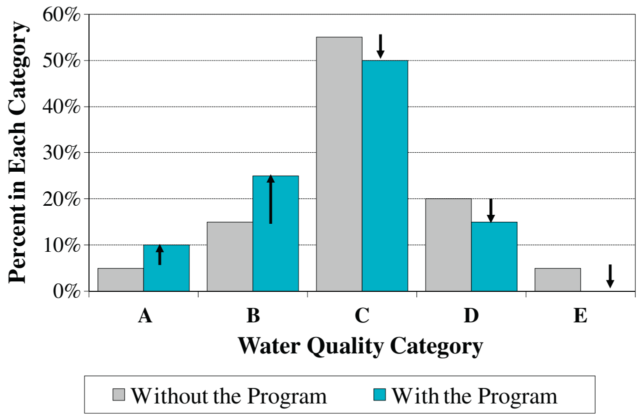

Fig. 3. Graphic used to show change in water quality.

[6] The KN panel has been effectively used in several economic studies, including Viscusi et al. (2008) for water quality. Cameron and DeShazo (2013) provide a more recent example; their technical appendix contains a careful discussion of the KN sample characteristics. Several studies have investigated the potential for nonresponse bias in stated preference surveys administered through KN. Studies that investigated various sources of nonresponse bias have found some evidence of sample selection in the demographic characteristics of the KN panel. However, the studies have found little evidence that the sample composition results in biased WTP estimates, and the differences that have been found were judged to be small (see Viscusi et al., 2008; Cameron and DeShazo, 2013).

[7] We did not use a conjoint choice experiment for this analysis because (1) the water quality attributes (color, clarity, etc.) do not vary independently in our design (for reasons previously discussed) and (2) the state-level percentages of lakes in each eutrophication category (A–E) also cannot vary independently from each other. In Phaneuf et al. (2013) we discuss a related choice experiment that was designed at the individual lake level.

Table 6
Summary statistics for SP survey data ($N = 1318$).

Variable name	Variable description	Mean	Std. Dev.	Min	Max
vote	=1 if respondent voted for program	0.566	0.496	0	1
uncertain	=1 if respondent is "very uncertain" about his/her vote	0.103	0.304	0	1
vote recode	=vote, except replaced with 0 if uncertain = 1	0.524	0.500	0	1
vote certain	=vote, except replaced with missing if uncertain = 1	0.585	0.493	0	1
bid	bid value presented to respondent ($ per year)	180.7	122.4	24	360
program II	=1 if respondent presented with program II	0.266	0.442	0	1
program III	=1 if respondent presented with program III	0.246	0.431	0	1
program IV	=1 if respondent presented with program IV	0.259	0.439	0	1
dquality	change in water quality index from baseline to with-program conditions (absolute value)	0.645	0.185	0.35	0.85
income	respondent's household income (10^4 2010 dollars)	6.285	4.506	0.25	20
college	=1 if respondent took at least one single day trip to a lake last year	0.333	0.471	0	1
triplastyr	=1 if respondent is a college graduate	0.324	0.468	0	1
tripnextyr	=1 if *triplastyr* = 0 but respondent is "somewhat" or "very" likely to take a single day lake trip next year	0.289	0.454	0	1
va	=1 if respondent lives in VA	0.165	0.371	0	1
nc	=1 if respondent lives in NC	0.275	0.447	0	1
Age	respondent's age	48.85	15.87	18	94
Female	=1 if respondent is female	0.561	0.496	0	1
Married	=1 if respondent is married	0.562	0.496	0	1
White	=1 if respondent is White	0.737	0.441	0	1

To elicit households' WTP we used a voting referendum dichotomous choice format, based in part on commonly recommended practices for CV studies (Arrow et al., 1993; Kling et al., 2012). The payment vehicle for the program was defined as a cost of living increase[8] using the following language:

> 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.

The bid amount *V* was randomly varied across respondent. Based in part on pretests of the survey instrument, we selected the following four annual bid values: $24, $120, $216, and $360. After responding to the bid question respondents were asked to report the level of certainty (i.e. very, somewhat, or not at all certain) that they felt about their answer.[9]

In April and May of 2010 KN initiated the survey by inviting 1873 adult members of their panel in our target states to complete the survey. KN recruits their panel of US households through random-digit dialing and, more recently, through address-based sampling.[10] In total, 1327 individuals completed the survey, resulting in a 70.8% completion rate. The full response rate for KN surveys is lower, due to nonresponse during panel recruitment and attrition rates among panel participants.

Table 6 displays summary statistics for the sample of respondents. Almost 60% of the sample was from the three main states – Virginia, North Carolina, and South Carolina – with the remainder roughly evenly divided between Alabama, Georgia, Kentucky, Mississippi, and Tennessee. As an indicator of familiarity with lake water quality we asked respondents if they had made a day trip (i.e. without an overnight stay) to a lake in the previous twelve months; 32% report making a lake visit. Among those who had not made a day trip in the previous year nearly 43% indicated they were 'somewhat' or 'very' like to do so in the following twelve months.

Overall, 57% of respondents indicated they would vote in favor of the lake water quality improvement program in their state. As expected, this percentage was inversely related to the annual payment. Among those presented with the $24 annual payment 73% voted to contribute; this fell to 42% for people presented with the $360 annual payment. This declining percentage of votes is consistent with expectations and provides suggestive evidence that the bid design (range and number of bids) was appropriate for estimating the marginal utility of income parameter (γ_1 below). The certainty follow up responses suggested that 40% were 'very certain', 50% were 'somewhat certain', and 10% were 'not certain at all'.

To analyze the responses to the referendum question, we use a utility difference framework as our conceptual model. We assume people 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 utility difference model for respondent *i* is

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

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, whose effect 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

$$\Pr(yes_i) = (1 + \exp(-\Delta V_i))^{-1}, \tag{2}$$

[8] A broad-based cost of living increase payment vehicle (rather than a targeted fee or tax) was selected in order to be most consistent with the broad range and distribution of pollution control measures included under the program. With this type of 'payment' it was also more realistic to define a permanent rather than a temporary increase in cost. Similar broad-based payment vehicles have successfully been applied in other stated preference studies (Boyle et al., 1994; Viscusi et al., 2008), and in our pretesting we did not find evidence of respondents rejecting this payment scenario.

[9] Our survey also contained the usual methods for limiting hypothetical bias and encouraging truthful responses. A brief cheap talk script (Cummings and Taylor, 1999) was included as a reminder of people's budget constraint, and consequentiality was stressed at various points in the survey.

[10] For more information on KN, see 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. The sample for any particular survey is randomly selected from KN's larger panel.

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

$$E\left(WTP_{ij}\right) = -\left(\alpha + \delta_j + \beta X_i\right)/\gamma_1. \quad (3)$$

An alternative way of modeling the role of statewide changes in the distribution of lake water quality categories is to combine the categories and percentages into a single percentage weighted average index as follows

$$qual_j = \left(1 \times p_j^A\right) + \left(2 \times p_j^B\right) + \left(3 \times p_j^C\right) + \left(4 \times p_j^D\right) + \left(5 \times p_j^E\right), \quad (4)$$

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. This approach not only simplifies the analysis, but it also imposes an assumed structure on preferences regarding the statewide distribution of lake water quality.

Using this assumption, the figures shown in Table 5 imply that the baseline average index is $qual_0 = 3.05$, while the improved average index values are $qual_I = 2.70$, $qual_{II} = 2.45$, $qual_{III} = 2.30$, and $qual_{IV} = 2.20$. Using this continuous average index for water quality our utility difference model becomes

$$\Delta U_i = \gamma_1 bid_i + \delta_1 \ln(\Delta qual_i + 1) + \delta_2 \ln(\Delta qual_i + 1) \times X_i + \alpha + \beta X_i + \varepsilon_i, \quad (5)$$

where $\Delta qual_i = qual_0 - qual_j$ is the improved quality level presented to respondent i and we have used a log transformation. Table 6 shows that the value of $\Delta qual_i$ ranges from 0.35 to 0.85, with a mean value of 0.645. For this model the expected WTP for a specific change in quality is

$$E(WTP_i) = -[\delta_1 \ln(\Delta qual + 1) + \delta_2 \ln(\Delta qual + 1) \times X_i + \gamma + \beta X_i]/\gamma_1. \quad (6)$$

Table 7 presents coefficient estimates for six different dichotomous choice logit models. The first three columns correspond to the program dummy variable specification in Eq. (1) and the last three columns correspond to the continuous quality difference specification in Eq. (5). 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'. 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 find evidence of hypothetical bias among uncertain respondents. Respondents who were certain of their responses showed little or no hypothetical bias. 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 $N = 1182$ 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. 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 characteristics, such as age, sex, race, and marital status were found to be individually and

Table 7
Logit regression analysis of CV survey responses.

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

Standard errors are in parentheses.
* $p < 0.10$.
** $p < 0.05$.
*** $p < 0.01$.

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. These results provide some, but not strongly significant, evidence of sensitivity to scope. We conclude from these positive estimates that the variability in quality levels amongst the programs was likely too small to detect statistically significant differences between *each* of them. However, the variability is large enough to detect some differences in preference between the programs with the smallest change (I) and those with the largest changes (III and IV). While our design could have varied the differences among programs to a larger degree, and thereby increased the statistical power to identify scope effects without functional form assumptions, we felt 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 Eq. (5), where the log transformation of the continuous quality attribute imposes a smooth diminishing marginal utility of the quality change.[11] 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.[12]

The results from all six models can be used to predict average willingness to pay for the water quality improvements. In addition, we estimate confidence intervals for each WTP estimate using the Krinsky and Robb (1986) simulation procedure. For example, using the formula in Eq. (3) and sample mean values for *college*, *triplastyr*, and *tripnextyr* 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.

As expected, recoding all uncertain votes to 'no' in model 2 leads to a lower mean WTP. Using the formula in Eq. (6) 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.

These welfare estimates, our summary statistics, and several auxiliary analyses provide evidence of the validity of our CV experiment (see Kling et al., 2012, for a recent summary of validity concepts). The probability of a yes vote falls when the cost to respondents increases, and scope effects are established qualitatively by the program dummy variable model and quantitatively by the continuous quality variable model. The confidence intervals for welfare measures from the two specifications overlap when similar improvements are considered; this lends support to the structure we have imposed on the continuous quality model. Finally, summary statistics for our certainty follow up question suggest respondents did not engage in 'yea saying'. Among the people who were uncertain about their answer, only 41% answered yes to the referendum, while among those who were very certain 58% answered yes. These validity checks suggest we can have some confidence in using our model for policy predictions. In the case study that follows we use model 5 from Table 7, because several studies (e.g. Blumenschein et al., 2008; Morrison and Brown, 2009) have shown that recoding uncertain responses as no votes provides estimates of WTP that are more likely to match estimates from real payment experiments.

4. Case Study — Benefits of Improvements in VA Lakes from the Chesapeake Bay TMDL

In this section we demonstrate how our protocol can be used to estimate the aggregate benefits of a specific statewide improvement in water quality. We use the Chesapeake Bay TMDL and its expected impacts on lake water quality in Virginia to conduct this demonstration. To meet its requirements under the Clean Water Act, the EPA established a TMDL for the Chesapeake Bay in 2010. This policy sets annual load limits on the amount of nitrogen, phosphorus, and sediment that may enter the Bay, with the objectives to be achieved by 2025. The watershed for the Chesapeake Bay includes portions of Delaware, Maryland, New York, Pennsylvania, Virginia, West Virginia, and the District of Columbia. These jurisdictions are responsible for developing Watershed Implementation Plans that specify how and where the load reductions will occur throughout the watershed. Although the federal and state efforts are specifically designed to improve conditions in the Bay, the resulting point and nonpoint source controls are also expected to reduce loads and improve water quality in upstream water bodies. We take advantage of the spatial overlap between our study area and the Bay watershed and focus on improvements to lakes located within the Virginia portion of the watershed. This spatial overlap is shown in Fig. 4.

4.1. Modeling Water Quality Improvement in VA Lakes

To predict how lakes in the Chesapeake Bay watershed will be affected by the TMDL, we used load reduction estimates from the Chesapeake Bay Community Watershed Model (USEPA, 2010) and applied them to the Northeast SPARROW model (Moore et al., 2011).[13] The load reduction runs were conducted by the Chesapeake Bay Program Office to simulate differences between pre-TMDL "baseline" conditions and those expected to prevail with the TMDL fully implemented; they provide estimates of the percentage change in total nitrogen and total phosphorus load to the sub-watersheds within the larger Bay watershed. These estimates describe changes in nutrient loads for streams in the Chesapeake drainage and the Bay itself, but not for lakes.

The Northeast SPARROW model estimates average annual nitrogen and phosphorus loads (kg/yr) to stream segments in the Mid-Atlantic and Northeast regions of the United States (including the Chesapeake drainage area) based on the long-term monitoring data and landscape

[11] We found a similar declining effect in a model using a quadratic specification for the water quality change; however, we selected the logarithmic form as our preferred specification because it eliminates utility decreases at higher levels of water quality changes.

[12] The models in Table 7 were also estimated with a two-stage Heckman sample selection model using demographic data on KN panel members who were invited to take the survey and declined. We did not find evidence of selection bias.

[13] Although focused on the northeastern regions of the U.S., the Northeast SPARROW model also includes the portion of Virginia that resides within the Chesapeake Bay watershed.

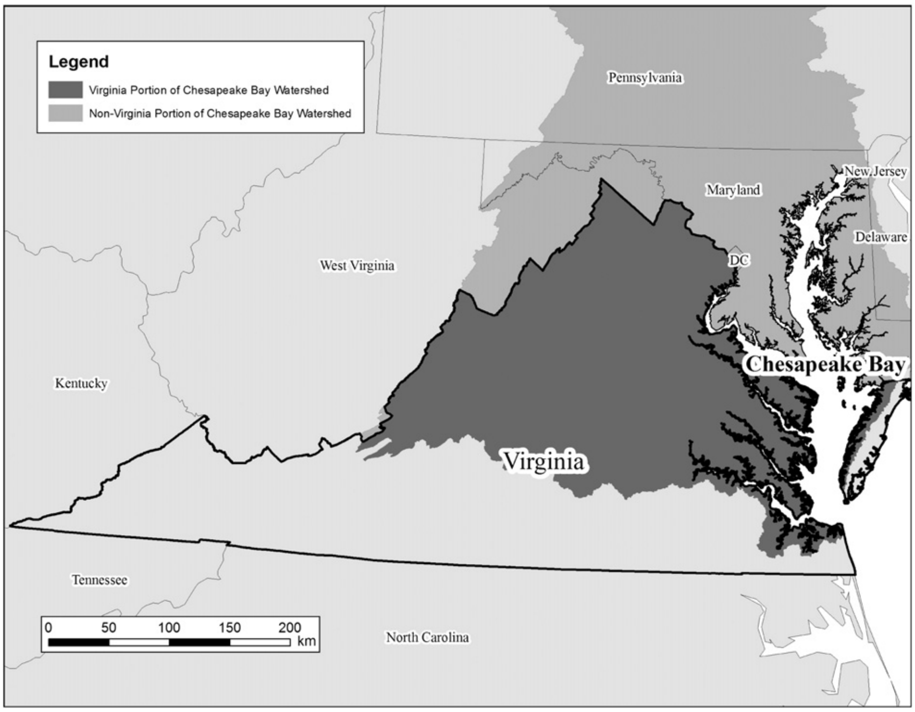

Fig. 4. Spatial overlap between Virginia and the Chesapeake Bay watershed.

conditions for the year 2002. It is based on NHDplus version 1 (National Hydrography Dataset) — a hydrologic network that includes representations of rivers, streams, lakes and other water bodies.[14] The Northeast SPARROW model is particularly useful for this analysis because its nutrient load predictions can be used to estimate concentrations of nitrogen and phosphorus in 2119 water bodies listed as lakes, ponds, and reservoirs (hereafter lakes) that are connected to the NHDplus network for the Virginia section of the Chesapeake watershed.[15] NHDplus also identifies 1208 lakes in Virginia that are outside the Chesapeake drainage; lakes within the watershed therefore account for roughly 64% of all lakes in Virginia.

Using SPARROW, we estimated nutrient loads to lakes on the NHDplus network as the sum of nutrient exports for all stream segments immediately upstream of a lake, plus the sum of incremental loads (loads originating within local catchments) for stream segments within the lake. We estimated water flow as the sum of water exports for all stream segments exiting the lakes. The nutrient concentrations for the lakes were then estimated by dividing the nitrogen and phosphorus loads by the water flows. Using this process we found that the predictions of long term average annual nutrient concentrations were consistent with, but higher than, the values for mid-summer nutrient concentrations observed via the 2007 National Lake Assessment (USEPA, 2009a,b) for lakes in the region (Milstead et al., 2013). The difference was due to inter-annual and seasonal variation in loads, and lower than expected estimates of nutrient retention in lakes from the SPARROW model.[16] If estimates of water residence time are available, Vollenweider equations predicting lake nutrient concentrations from input load and residence time can be used to more accurately estimate nutrient concentrations in lakes (Reckhow, 1988; Vollenweider, 1975). We estimated lake retention times from modeled lake volumes (Hollister and Milstead, 2010; Hollister et al., 2011) and flow predictions from SPARROW. These estimates, along with predicted nutrient loads and observed nutrient concentration from the National Lake Assessment, were used to develop and estimate Vollenweider equations following the methods of Milstead et al. (2013), from which nitrogen and phosphorus concentrations in the lakes were predicted.

To understand how the Chesapeake Bay TMDL will affect nutrient concentration in lakes, we used the US Geological Service SPARROW decision support system, an online platform that facilitates querying and scenario analysis of the Northeast and other SPARROW models (Booth et al., 2011). To link the Chesapeake Bay Community Watershed Model load reduction estimates to SPARROW, all NHDplus stream segments in the Chesapeake drainage were assigned sub-watersheds, using a GIS intersect procedure.[17] The percent changes in total nitrogen and phosphorus predicted for the NHDplus stream segments were submitted to the SPARROW decision support system for analysis. The system returned predicted nitrogen and phosphorus loads for stream segments based on the TMDL scenario, and these estimates were translated into lake nutrient concentrations as described above.

The final step was to estimate chlorophyll *a* concentrations. Using National Lake Assessment observations from 29 lakes in the Chesapeake drainage, we developed a linear model of the relationship between nitrogen, phosphorus and chlorophyll *a* concentrations. The resulting regression equation

$$\log_{10}(chla) = -0.48 + 0.88 \times \log_{10}(TP) + 0.09 \times \log_{10}(TN) \tag{7}$$

was used to estimate Chlorophyll *a* concentrations for lakes in the Chesapeake drainage under the baseline and TMDL scenarios.

[14] See www.horizon-systems.com/NHDPlus/NHDPlusV1_home.php for details.

[15] Most of the other 970 lakes that do not have SPARROW predictions are salt ponds, ephemeral ponds, or isolated systems, which are not connected to the stream network and therefore should not be included in this analysis.

[16] Nutrient retention refers to the portion of nutrients entering a lake that accumulate in the sediments, are consumed by plants and animals, or are lost to the system by chemical processing. In lakes, nutrient retention depends largely on the water residence time (flow/volume). The SPARROW model derives an estimate of nutrient retention from the data, but these estimates tend to be low for lakes with long residence times and high for those with short residence times.

[17] Stream segments that intersected more than one sub-watershed were assigned to the one that contained the largest proportion of its linear extent.

Table 8
Modeled water quality for VA lakes in Chesapeake Bay watershed ($N = 2119$).

	Baseline	With TMDL
Mean total nitrogen (mg/l)	0.881	0.719
Mean total phosphorus (mg/l)	0.084	0.071
Mean chlorophyll a (μg/l)	36.1	30.4
Percent by trophic category/index		
A/1	2%	4%
B/2	14%	19%
C/3	42%	46%
D/4	26%	22%
E/5	17%	10%
Average trophic index (*qual*)	3.41	3.15

Table 8 summarizes the modeling results for the 2119 lakes located in the Virginia potion of the Chesapeake Bay watershed. Under baseline conditions, average nitrogen and phosphorus levels are 0.88 and 0.08 mg/l respectively, and chlorophyll *a* is 36.10 μg/l. With the TMDL, these levels decrease by 16 to 18%.

4.2. Change in Trophic Distribution of VA Lakes

To translate these pollutant concentration estimates into trophic categories, we apply the results from the expert elicitation analysis. In particular, we use the ordered logit results from model 3 in Table 3 to predict the trophic category/index for each of the 2119 lakes under baseline and with-TMDL conditions. Although model 2 in Table 2 provides a better statistical fit than model 3, it requires data on turbidity which are not available from our water quality model. We turn to model 3 as the best alternative, given our data limitations. Table 8 reports the resulting frequency distributions. Moving from the baseline to TMDL policy regimes causes the percentage of lakes in the lowest quality categories (D/4 and E/5) to decrease from 43 to 32%, and those in the highest two categories (A/1 and B/2) increases from 16 to 23%. Using Eq. (4), we can convert these frequency distributions into a percentage-weighted average index (*qual*), which is 3.41 in the baseline and 3.15 with the TMDL. Therefore, the average trophic index level changes by 0.26 points (on the 1 to 5 scale).

For all the lakes in Virginia, including those outside the Chesapeake Bay watershed, the shift in water quality distribution is somewhat different. Water quality in the 36% of Virginia lakes that are located outside of the Bay watershed is not expected to changes as a result of the TMDL. Factoring in these additional lakes, we estimate the change in the average trophic index level for *all* lakes in the state due to the TMDL to be $\Delta qual = 0.16$ index points.

4.3. Aggregate Benefit Estimates for VA Households

In the final step we use the results from the contingent valuation model to the estimate average WTP among Virginia households for the water quality change, and we multiply this value by the total number of households in Virginia to estimate aggregate benefits for the state's residents. Specifically, we apply Eq. (6) with the parameter estimates from model 5 in Table 7. 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. To specify the vector of household characteristics – *college*, *triplastyr*, and *tripnextyr* – we use the mean values (0.392, 0.263, and 0.277 respectively) from our sample of survey respondents from Virginia ($N = 217$). The resulting mean annual WTP estimate is $60 per household, with a 95% confidence interval of ($51, $70). Aggregating across 3.06 million households (based on 2010 census data) in the state, the total estimated benefits for Virginia households from lake water quality improvements going from baseline conditions to TMDL conditions are $184 million per year in 2010 dollars, with a 95% confidence internal of ($157 million, $214 million).[18]

5. Conclusion

As states and other jurisdictions continue to develop numeric nutrient criteria and load limits to protect surface waters, it is important for policymakers to have analytical tools that allow them to gauge the ecosystem service benefits resulting from these rules. In a recent *PNAS* paper Keeler et al. (2012) discuss the extent to which current economic and ecological modeling approaches provide these tools. They write that:

> "…most water quality valuation tools fall short of the needs and expectations of decision makers. [One] shortcoming… is that valuation assessments are often not linked with changes in management … that lead to water quality changes…. Finally, economic models for valuing water quality related ecosystem services are often poorly integrated with ecological and hydrological models."
>
> [pp. 18619–18620]

Keeler et al. go on to suggest a template for integrating the various models and data in a way that would address these limitations. Our paper describes an operational version of this template and illustrates how it can meet the needs of policy makers while avoiding many of the limitations of past studies. Most importantly, the framework provides a link between changes in nutrient related water and human welfare by integrating environmental assessment and economic valuation methods. Our innovation is to use expert elicitation to estimate a function that translates multiple nutrient-related water quality parameters into on ordinal index of water quality categories, and a parallel stated preference survey that maps the index levels to observable features of water bodies using lay audience descriptions. The valuation application then uses changes in index levels – expressed as a shift in the distribution of water quality at lakes in the state – as the commodity to be valued. By closely coordinating the outputs from our environmental modeling with the inputs needed for economic modeling we provide a protocol that allows an analyst to directly trace the effects of a policy from changes in nutrient loads to changes in numeric ecosystem health indicators, and on to ecosystem services and values.

We use the Virginia TMDL case study to demonstrate how our general framework can be applied in specific application. It also serves to demonstrate the advantages and main contributions of this approach compared to other existing approaches. For example, existing methods can be used to translate multiple water quality parameters into a water quality index (Cude, 2001; Vaughn, 1986) and benefit transfer functions based on water quality indexes also exist (Carson and Mitchell, 1993; Johnston et al., 2005; Van Houtven et al., 2007); however, there is typically little evidence to demonstrate the correspondence between the indexes used in these two functions. In contrast, our approach is specifically designed to ensure correspondence. Moreover, our approach focuses specifically on nutrients and eutrophication in lakes, whereas other approaches often add in other pollutants and water bodies, which further complicate the use and interpretation of a single index. Finally, our approach defines the spatial boundaries of water quality changes (i.e., in-state) and incorporates the distribution (percentage) of lakes across multiple quality categories. These combined

[18] We caution against interpreting these results as the benefits specifically *attributable* to the Chesapeake Bay TMDL, in particular because some portion of the expected load reductions and water quality improvements by 2025 would most likely have occurred even without the TMDL. Instead, the estimates represent the benefits of the difference between before-TMDL and after-TMDL conditions.

features are most often not included in other existing benefit transfer functions for water quality.

Despite these advantages, it is also important to recognize the limitations of our approach and to interpret the results of the case study accordingly. In particular, the calculation and use of an average index (*qual* from Eq. (4)) to combine the statewide distribution of lake trophic water quality into a single continuous measure is convenient but requires somewhat strong simplifying assumptions. Most importantly it interprets a 5-level ordinal index as a cardinal measure of water quality and assumes that the average measure across lakes is what matters to individuals. Although the change in this average index (in logarithmic form) is statistically significant in our regression analyses of the contingent valuation data, it is not necessarily the most accurate representation of preferences. Nevertheless, other benefit transfer approaches applied to state or nationwide changes in water quality (e.g., USEPA, 2000, 2009a,b) have required similar simplifying assumptions.

Our sense is that coordinated expert elicitation and stated preference modeling has great potential for use in analysis of environmental policy. The water quality ideas presented here could be replicated for other regions, and comparisons made amongst similar model structures calibrated to the varying spatial conditions. More generally, policy decisions related to, for example, air quality and biodiversity preservation could benefit from our approach. As an air quality example, it would be useful to quantify experts' opinions on how ambient levels of particulate matter map to the severity of illness or duration of symptoms for people of varying ages and existing health status. These could in turn be the commodities that people express preferences for in a stated preference survey. As a biodiversity example, quantifiable opinions on how different sizes and positioning of habitat preservation map to actual improvements in survivability of specific species would be a useful way to connect things over which people have preferences for – species protection – with the policy lever available to policy makers.

Implementing our protocol for water quality required several steps that condensed complex systems and processes into manageable dimensions. This would undoubtedly be the case for other applications of our protocol, and so research should focus on explicitly identifying the consequences of any needed simplifications. An example in our case corresponds to the linear nature of our quality index. The five levels of lake water quality in our survey (summarized in Table 4) were designed to correspond to the rankings used by the experts. As previously discussed, based in part on our focus group work we decided *not* to allow the individual attribute levels (color, clarity, fish, etc.) to vary independently, in order to minimize respondents' cognitive burden and the potential for invalidity that the added complexity would imply. This decision means we are not able to consider the potential for nonlinear relationships between attribute levels and the index rankings. Consideration of the consequences for policy analysis of these types of decisions should be part of any subsequent research on this method.

References

Arrow, Kenneth, Solow, Robert, Portney, Paul, Leamer, Edward E., Radner, Roy, Schuman, Howard, January 15 1993. Report of the NOAA panel on contingent valuation. Fed. Regist. 58 (10), 4602–4614.

Banzhaf, H. Spencer, Burtraw, Dallas, Evans, David, Krupnick, Alan, 2006. Valuation of natural resource improvements in the Adirondacks. Land Econ. 82, 445–464. http://dx.doi.org/10.3368/le.82.3.445.

Blumenschein, K., Blomquist, G.C., Johannesson, M., Horn, N., Freeman, P., 2008. Eliciting willingness to pay without bias: evidence from a field experiment. Econ. J. 118, 114–137.

Booth, N.L., Everman, E.J., Kuo, I.L., Murphy, L., Sprague, L., 2011. A web-based decision support system for assessing regional water-quality conditions and management actions. J. Am. Water Resour. Assoc. 47 (5), 1136–1150.

Boyle, Kevin J., Desvousges, William H., Reed Johnson, F., Dunford, Richard W., Hudson, Sara P., 1994. An investigation of part–whole biases in contingent-valuation studies. J. Environ. Econ. Manag. 27 (1), 64–83.

Cameron, T.A., DeShazo, J.R., 2013. Demand for health risk reductions. J. Environ. Econ. Manag. 65, 87–109.

Carson, R.T., Mitchell, R.C., 1993. The value of clean water: the public's willingness to pay for boatable, fishable, and swimmable quality water. Water Resour. Res. 29 (7), 2445–2454.

Champ, P.A., Bishop, R.C., Brown, T.C., McCollum, D.W., 1997. Using donation mechanisms to value nonuse benefits from public goods. J. Environ. Econ. Manag. 33 (2), 151–162.

Cude, C.G., 2001. Oregon water quality index: a tool for evaluating water quality management effectiveness. JAWRA J. Am.Water Resour. Assoc. 37 (1), 125–137.

Cummings, R.G., Taylor, L.O., 1999. Unbiased value estimates for environmental goods: a cheap talk design for contingent valuation. Am. Econ. Rev. 89, 649–665.

Edwards, S., Anderson, G., 1987. Overlooked biases in contingent valuation surveys: some considerations. Land Econ. 63, 168–178.

Egan, K.J., Herriges, J.A., Kling, C.L., Downing, J.A., 2009. Valuing water quality as a function of water quality measures. Am. J. Agric. Econ. 91, 106–123.

Herriges, J., Kling, C., Liu, C.C., Tobias, J., 2010. What are the consequences of consequentiality? J. Environ. Econ. Manag. 59 (1), 67–81.

Hollister, J., Milstead, W.B., 2010. Using GIS to estimate lake volume from limited data. Lake Reservoir Manage. 26 (3), 194–199.

Hollister, J.W., Milstead, W.B., Urrutia, M.A., 2011. Predicting maximum lake depth from surrounding topography. PLoS ONE 6 (9), e25764.

Johnston, R.J., Besedin, E.Y., Iovanna, R., Miller, C.J., Wardwell, R.F., Ranson, M.H., 2005. Systematic variation in willingness to pay for aquatic resource improvements and implications for benefit transfer: a meta-analysis. Can. J. Agric. Econ. 53, 221–248.

Keeler, B.L., Polasky, S., Brauman, K.A., Johnson, K.A., Finlay, J.C., O'Neill, A., Kovacs, K., Dalzell, B., 2012. Linking water quality and well-being for improved assessment and valuation of ecosystem services. Proc. Natl. Acad. Sci. 109 (45), 18619–18624.

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.

Kenney, M.A., Arhonditsis, G.B., Reiter, L.C., Barkley, M., Reckhow, K.H., 2009. Using structural equation modeling and expert elicitation to select nutrient criteria variables for south-central Florida lakes. Lake Reservoir Manage. 25 (2), 119–130.

Kling, C.L., Phaneuf, D.J., Zhao, J., 2012. From Exxon to BP: has some number become better than no number? J. Econ. Perspect. 26, 3–26.

Krinsky, I., Robb, A.L., 1986. On approximating the statistical properties of elasticities. Rev. Econ. Stat. 68, 715–719.

McClelland, N.I., 1974. Water Quality Index Application in the Kansas River Basin. (EPA-907/9–74–001) U.S. EPA Region VII, Kansas City, MO.

Milstead, W.B., Hollister, J.W., Moore, R.B., Walker, H.W., 2013. Estimating nutrient concentrations in Northeastern lakes from predicted loads and modeled lake depths (in preparation).

Mitchell, R.C., Carson, R.T., 1984. Willingness to Pay for National Freshwater Quality Improvements. Draft Final Report to the U.S. EPA (CR 810224–01–4).

Moore, R.B., Johnston, C.M., Smith, R.A., Milstead, B., 2011. Source and delivery of nutrients to receiving waters in the northeastern and mid-Atlantic regions of the United States. J. Am. Water Resour. Assoc. 47 (5), 965–990.

Morgan, M.G., Henrion, M., 1990. Uncertainty: A Guide to Dealing with Uncertainty in Quantitative Risk and Policy Analysis. In: Cambridge University Press (Ed.), Cambridge University Press.

Morrison, M., Brown, T., 2009. Testing the effectiveness of certainty scales, cheap talk, and dissonance-minimization in reducing hypothetical bias in contingent valuation studies. Environ. Resour. Econ. 44, 307–326.

Phaneuf, D.J., Kenney, M.A., Reckhow, K.H., May, 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 to U.S. EPA. U.S. EPA Office of Water, Washington, D.C. (Grant #X7-83381001-0).

Phaneuf, D.J., von Haefen, R.H., Mansfield, C.A., Van Houtven, G., February, 2013. Measuring Nutrient Reduction Benefits for Policy Analysis Using Linked Non-market Valuation and Environmental Assessment Models. Final Report on Stated Preference Surveys. Report to U.S. EPA. U.S. EPA Office of Water, Washington, D.C (Grant #X7–83381001–0).

Reckhow, K.H., 1988. Empirical models for trophic state in Southeastern U.S. lakes and reservoirs. J. Am. Water Resour. Assoc. 24 (4), 723–734.

Reckhow, K.H., Arhonditsis, G.B., Kenney, M.A., Hauser, L., Tribo, J., Wu, C., Steinberg, L.J., Stow, C.A., McBride, S.J., 2005. A predictive approach to nutrient criteria. Environ. Sci. Technol. 39 (9), 2913–2919.

Roberts, D.C., Boyers, T.A., Lusk, J.L., 2008. Preferences for environmental quality under uncertainty. Ecol. Econ. 66, 584–593.

Smith, V.K., Desvousges, William H., 1986. Measuring Water Quality Benefits. Kluwer-Nijhoff, Boston.

Sonneveld, B.G.J.S., Albersen, P.J., 1999. Water erosion assessment based on expert knowledge and limited information using an ordered logit model. J. Soil Water Conserv. 54 (3), 592–599.

USEPA, December 2000. Ambient Water Quality Criteria Recommendations: Information Supporting the Development of State and Tribal Nutrient Criteria: Lakes and Reservoirs in Nutrient Ecoregion IX. (EPA-822-B-00-011) U.S. EPA, Office of Water, Washington, DC.

USEPA, December 2002. Estimation of National Economic Benefits Using the National Water Pollution Control Assessment Model to Evaluate Regulatory Options for Concentrated Animal Feeding Operations. (EPA-821-R-03-009) U.S. EPA, Office of Water, Washington, DC.

USEPA, 2009a. National Lakes Assessment: A Collaborative Survey of the Nation's Lakes. Washington, D.C. EPA 841-R-09-001. http://water.epa.gov/type/lakes/upload/nla_newlowres_fullrpt.pdf.

USEPA, 2009b. Environmental Impact and Benefits Assessment for Final Effluent Guidelines and Standards for the Construction and Development Category. U.S.

Environmental Protection Agency, Office of Water, Washington, DC (EPA-821-R-09-012).
USEPA, December 2010. Chesapeake Bay Phase 5.3 Community Watershed Model. (EPA 903S10002-CBP/TRS-303–10) U.S. Environmental Protection Agency, Chesapeake Bay Program Office, Annapolis MD.
Van Houtven, G.L., Powers, J., Pattanayak, S.K., 2007. Valuing water quality improvements using meta-analysis: is the glass half-full or half-empty for national policy analysis? Resour. Energy Econ. 29, 206–228.
Vaughn, W.J., 1986. "The Water Quality Ladder". Appendix to RC Mitchell and RT Carson, A Contingent Valuation Estimate of National Freshwater Benefits. Prepared for the US Environmental Protection Agency, Resources for the Future.
Viscusi, W.K., Huber, J., Bell, J., 2008. The economic value of water quality. Environ. Resour. Econ. 41, 169–187.
Vollenweider, R., 1975. Input–output models. Aquatic Sciences — Research Across Boundaries 37, 53–84.