PNAS PERSPECTIVE

Perspectives on valuing water quality improvements using stated preference methods

Ian J. Bateman[a,1], Bonnie Keeler[b], Sheila M. Olmstead[c], and John Whitehead[d]

Edited by David Zilberman, University of California Berkeley, Berkeley, CA; received October 13, 2022; accepted December 21, 2022

Improvements to the quality of freshwater rivers and lakes can generate a wide array of benefits, from "use values" such as recreational boating, fishing, and swimming to "nonuse values" such as improved outcomes for aquatic biodiversity. Bringing these nonmarket values into decision-making is crucial to determining appropriate levels of investment in water quality improvements. However, progress in the economic valuation of water quality benefits has lagged similar efforts to value air quality benefits, with implications for water policy. New data sources, modeling techniques, and innovation in stated preference survey methods offer notable improvements to estimates of use and nonuse benefits of improved water quality. Here, we provide a perspective on how recent applications of stated preference techniques to the valuation of the nonmarket benefits of water quality improvements have advanced the field of environmental valuation. This overview is structured around four key questions: i) What is it about water quality that we seek to value? ii) How should we design and implement the surveys which elicit individuals' stated preferences? iii) How do we assess the validity of the findings provided by such studies? and iv) What are the contributions of these valuation exercises to public policy? In answering these questions, we make reference to the contributions provided by the papers in this Symposium.

economics | valuation | water quality | stated preference

As rules promulgated under the major US environmental statutes of the 1970s and earlier have been applied more broadly, environmental economists have developed and improved environmental benefit valuation techniques to explore the associated tradeoffs. This exercise is not purely academic. An important driver of the development of such methods has been a set of US Presidential executive orders dating from the 1980s, which require rigorous benefit-cost analysis for "economically significant rules"—regulations with an anticipated economic cost exceeding $100 million per year (1). A large share of these economically significant rules has been applied by the US Environmental Protection Agency (EPA), the regulatory agency whose mission is to protect human health and safeguard the natural environment. Thus, benefit valuation techniques developed by economists play an important and visible role in making and evaluating environmental policy.

Despite decades of work, notable omissions still remain in the quantification of water quality benefits. These omissions limit the accuracy of EPA to quantify the benefits of prospective regulations. Sinden (2) examined 45 analyses of EPA and found that in 80% of those documents "... EPA excluded categories of benefits that the agency itself described as either actually or potentially 'important,' 'significant,' or 'substantial' because they were unquantifiable due to data limitations." Petrolia et al. (3) confirmed this finding for an additional set of analyses. So, while benefit techniques capable of valuing some of EPA's needed endpoints exist, the empirical studies needed to support these regulatory impact assessments do not. As a consequence of these gaps, EPA recently funded a set of innovative water quality valuation studies; findings and results of this work are the topic of this Symposium.

In this Perspectives piece, we review how this collection of papers extends the research frontier, yet leaves open several important questions for future research. These studies are united not only in terms of the focus of their valuations but in a number of key methodological challenges. In the remainder of this paper we discuss three issues common to the empirical studies presented: i) defining the public good under evaluation; ii) designing and implementing the surveys used in all these applications; and iii) assessing the validity of findings. Finally, we discuss the overall contributions to public policy provided by these studies.

Defining the Good: What Is It about Water Quality That We Value?

Changes in water quality affect a diverse suite of endpoints and an even greater set of human uses including bird watching, boating, duck hunting, fishing, picnicking, swimming, nongame wildlife viewing, and cultural values. In addition to these use values, people may also value clean water due to the role water quality plays in supporting nonuse values such as aquatic biodiversity.

The science linking water policies to changes in water quality is also complex—requiring assumptions about how activities such as application of fertilizer, stormwater management, or stream restoration affect multiple pollutants at varying spatial and temporal scales. Progress in integrated

Author affiliations: [a]Land, Environment, Economics and Policy Institute, Department of Economics, University of Exeter Business School, Exeter, EX4 4PU, United Kingdom; [b]Center for Science, Technology and Environmental Policy, Hubert H. Humphrey School of Public Affairs, University of Minnesota, Minneapolis, MN 55455; [c]Lyndon B. Johnson School of Public Affairs, The University of Texas at Austin, Austin, TX 78713-8925; and [d]Department of Economics, Appalachian State University, Boone, NC 28608

Author contributions: I.J.B., B.K., S.M.O., and J.W. wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission.

[1]To whom correspondence may be addressed. Email: i.bateman@exeter.ac.uk.

Published April 24, 2023.

assessment models, remote sensing, and real-time water quality monitoring has greatly improved our ability to predict how water quality is affected by human activities (4, 5). However, the science and practice of water valuation remains an important and active area of research due to the dynamic nature of the biophysical, social, and economic factors that determine water values and the persistence of water quality problems.

Given the complexity of water quality and the upstream and downstream connections to human values and uses, the choice of water quality metric—or the commodity to be valued—can be quite important. Ideally, water quality metrics used in economic valuation should be

- sensitive to the policy or program to be evaluated
- easy to interpret by respondents
- observable by the target population whose values will be elicited
- measurable in biophysical terms that relate to hydrologic or ecological processes
- commensurable, such that a change can be measured in monetary terms
- locally relevant yet generalizable to national scales appropriate for federal policy evaluation.

Water quality can be described in biophysical terms (e.g., concentrations of nitrogen), as a single water quality "score" based on a combination of factors or as a set of multiple qualitative or quantitative metrics. Simplified metrics reduce cognitive burden on respondents but may miss important aspects of water quality. Biophysically based metrics can be modeled or measured using existing data but may not be relevant or observable to respondents or tied to aspects of water quality that people value (6). An additional challenge is to develop water quality metrics that have the potential to quantify both use and nonuse values.

The assembled papers build on decades of work by economists and water quality scientists to develop metrics of water quality designed to capture a diversity of water values. Early work by McClelland (7) and Mitchell and Carson (8) used expert opinion to link measurable water quality parameters (e.g., fecal coliform, total nitrogen, and dissolved oxygen) with designated uses (boating, rough fishing, game fishing, swimming, and drinking). Valuation studies could then be structured to elicit respondents' willingness to pay for an increase in water quality based on comparisons between baseline and alternative scores on the water quality ladder (WQL) originally developed by McClelland (7). Consistent use of the WQL over time and space means that there is a large bank of past studies for comparison and to aid in national analyses.

At the same time, the assembled papers seek to improve upon several limitations of the WQL. Presented as discreet "rungs," the WQL does not allow for the assessment of marginal changes in water quality that do not cross defined thresholds for specific designated uses. These designated uses describe a limited set of water quality benefits and omit ecosystem or cultural services outside boating, fishing, swimming, or drinking. Last, the WQL simplifies complex ecological interactions that do not apply consistently to all waters. The assembled papers offer alternative metrics of water quality for use in economic valuation and provide useful insights into the application and interpretation of each metric.

In Lupi et al. (9) (in volume), three alternatives to the WQL metric are applied to a valuation survey of Michigan residents. A Fishing Biomass Score is used to capture game fishing conditions, a Water Contact Score is developed for nonfishing recreation such as swimming, kayaking, and boating, and an index of biological condition is used to represent the nonuse value of aquatic wildlife. Respondents were presented with alternative policy scenarios associated with changes in the spatial distribution of individual and combined water quality scores across the study region.

In von Haefen et al. (10) (in volume), six water quality parameters (biotic integrity, fecal coliform, specific conductance, total nitrogen, total phosphorus, and turbidity) were combined into metrics of "ecosystem condition," "human health risk," and "murky water days." Changes in water quality were presented as the percentage of stream miles rated as "high, med, or low" or "poor, fair, or good" in each of the three metrics.

Vossler et al. (11) (in volume) adopted a water quality measure developed by the EPA, the Biological Condition Gradient (BCG), that captures the diversity and relative abundance of freshwater taxa associated with a specific waterbody. The BCG metric is designed to reflect departures in ecological integrity from a reference or "undisturbed" condition. The authors convey changes in BCG via illustrations of biological diversity, visual conditions representing changes in water quality, and associated survey narratives designed to connect aquatic ecosystem health with attributes people can perceive, understand, and thus value.

The work of Hill et al. (12) (in volume) aimed to develop a nationally representative metric of aquatic biodiversity that could be used to elicit nonuse values for improved water quality. Designed to complement the WQL, the authors tested several potential metrics designed to capture "biological integrity." The authors settled on a metric of benthic macroinvertebrates based on data from the National Rivers and Streams and National Lakes Assessments and identify ways the metric can inform future stated preference surveys for nonuse values.

Johnston et al. (13) (in volume) assess the water quality benefits of riparian buffers, wastewater treatment plants, stormwater retention, and road salt application on rivers and streams in Connecticut, Massachusetts, Rhode Island, Vermont, New Hampshire, and Maine. The authors presented respondents with changes in three water quality metrics—water safety based on fecal coliform concentrations, support for aquatic life based on effects of chloride concentrations, and water pollution based on modeled solute concentrations. The survey communicated each metric in three different ways: a normalized, spatial mean value (0 to 100) over the river system, a bar chart showing the proportion of total river and stream miles within seven binned quality intervals, and geographical information system (GIS) maps showing water quality predictions over the river system.

Of the five papers, four included at least one metric of biological condition that could be considered a proxy for nonuse value. All papers, with the exception of Hill et al. (12), included use-based metrics of water quality in their assessments, either by associating changes in biological condition with changes in designated uses or by including a separate metric that captured changes in human use of aquatic resources. The assembled papers provide researchers with a suite of potential water quality metrics that can inform future valuation studies aiming to capture nonuse values for water quality.

We know from Lupi et al. (9) that elicited values are sensitive to the presentation of water quality metrics separately, rather than jointly. Policies that affect water quality may cause alternative metrics to move in opposite directions, such as a program that increases water clarity at the expense of desired game species. Lupi et al. (9) illustrate the potential for aggregated metrics such as the WQL to undervalue water quality benefits when multiple indicators are combined. All papers illustrate how the selection of water quality metrics narrows or expands the suite of values to be valued and poses risks of either undervaluing or double counting values, with implications for benefits transfer and resulting policy analyses.

von Haefen et al. (10) and Johnston et al. (13) suggest that not all water quality metrics are valued equally. von Haefen et al. (10) found that of the six water quality indicators evaluated, the percentage of high human health risk stream miles had the largest effect on the probability of voting for a hypothetical action plan. Similarly, Johnston et al. (13) find the greatest willingness to pay for "water safety," with improvements in total water pollution and aquatic life associated with positive but lower willingness to pay (WTP). Vossler et al. (11) suggest that nonuse sources of value (e.g., bequest value, intrinsic aesthetic value) are an important part of overall water quality benefits, but do not estimate the proportion of total value attributed to use vs. nonuse value.

No single metric for water quality will be comprehensive across all values and uses nor will all metrics be appropriate for all contexts. Tailored sets of local metrics based on observed biophysical data provide advantages for local policy evaluation but may limit the ability to scale study insights into other geographic regions or contexts. Vossler et al. (11) and Hill et al. (12) present water quality metrics that can facilitate national-scale evaluation, whereas Lupi et al. (9), von Haefen et al. (10), and Johnston et al. (13) opt for water quality metrics that are relevant to specific geographies and target populations. A notable gap across all papers is the explicit consideration of nonmaterial or relational values for clean water such as spiritual or cultural values that might influence willingness to pay for specific water bodies.

The relationship between water quality metrics and the diversity of water uses and values remains a challenge. The science of water quality is complex and constantly evolving, revealing new understandings of how interactions among constituents and across ecological gradients shape water quality outcomes. The papers illustrate how collaboration with water quality scientists via integrated models, integration of new data sets, and the use of expert elicitation can improve the accuracy of water quality valuation.

Designing and Implementing the Surveys

Early efforts using stated preference methods to estimate the values for changes in the allocation of environmental and natural resources involved simple bidding games and open-ended willingness to pay questions. These have evolved to discrete choice valuation questions with variation over costs and other attributes of the choice (e.g., environmental quality) with sophisticated experimental designs. The stated preference articles in this Symposium represent the state-of-the-art, especially considering the integration of environmental science and measurement of environmental attributes into the valuation exercise. This section of this Perspectives piece focuses on decisions made about valuation methods. One conclusion is the striking differences in form that these studies take.

Valuation Question Format. Since the landmark work by Carson and Groves (14), researchers who have employed the contingent valuation method have tended to employ a single binary choice valuation question with a coercive payment vehicle (e.g., a referendum vote with an unavoidable tax payment). For example, a question which asks a survey respondent if they would vote in favor of a policy that generates a specified environmental quality improvement and accompanied by a specified increase in taxes. The survey is considered to be consequential if survey respondents think its results could affect policy and in that sense represents "revealed" economic behavior in the same sense as voting. This particular format (with some auxiliary conditions) can be shown to be incentive compatible in the sense that the respondent should only indicate they favor the policy change if their willingness to pay for the environmental improvement is greater than the specified cost.

A single binary discrete question is a special case of a general class of preference elicitation models known as discrete choice experiment (DCE) that randomly assign respondents to choice alternatives that differ by one or more attribute levels (e.g., price and water quality levels). More general DCEs involve either offering respondents more than one alternative to the current status quo, ask respondents more than one choice question or both. These more general DCEs can provide considerably more information about the public's preferences than a single binary discrete choice question. However, as Carson and Groves (15) point out, collection of this extra preference information results in effects on the magnitude of WTP estimates that are not well known, although their direction, particularly with additional assumptions with strong empirical support, is discernable.

Two of the studies, Lupi et al. (9), and Johnston et al. (13) employ the single binary choice question format. The other two stated preference studies, von Haefen et al. (10) and Vossler et al. (11), ask multiple binary choice valuation questions. The key assumption needed for incentive compatibility is that respondents treat each choice question independently rather than engage in strategic behavior across questions to influence the implemented option. With the respondent asked to choose between multiple options in the same question, the issue of "how" a single alternative will be picked by the government emerges when only one can be implemented at any single point in time. von Haefen et al. (10) and Vossler et al. (11) employ four and (an average of) nine questions, respectively. Vossler et al. (11) explicitly ask respondents to treat each question independently, while von Haefen et al. (10) test whether responses to the first valuation question are different from responses to the other questions and find no statistical difference. One should, of course, not draw definitive inference about the performance of a particular preference elicitation format from a small set of studies focused on a particular type of environmental amenity. Nevertheless, these results are encouraging. The repeated referendum DCE-type stated that the preference valuation question may emerge as the dominant question format since it appears to achieve a reasonable degree of incentive compatibility while increasing statistical efficiency relative to the single binary choice question. If so, a key research issue is what is the "optimal" number of choice questions? This issue arises because it is known that boredom/fatigue can set in with repeated asking of choice questions.

Experimental Design. The simplest experimental design for a binary choice valuation question is for each respondent to be presented with a randomly assigned cost amount while the alternative to the status quo is defined by a fixed set of attributes (quantitative or qualitative) relative to that status quo. Clearly, the levels of these other attributes can also be randomly assigned to generate a "scope test" (16–18) revealing the responsiveness of WTP estimates to changes in the magnitude of the good under investigation. The experimental designs in this Symposium are considerably more complicated. Each of the studies presents three noncost attributes to survey respondents that reflect different water quality endpoints and each are communicated in different ways. Similarly, each of the four studies includes a cost attribute that differs in the payment vehicle (increases in taxes, fees, utility bills), payment frequency (monthly, annually), and payment schedule (one-time, annually for 5 y, in perpetuity).

Vossler et al. (11) consider regional improvements in water quality in the Midwest. They vary the spatial scale of the water quality improvement, the magnitude of the water quality improvement and whether the respondent lives in the geographic area where the water quality improvement would occur. The levels of the attributes are communicated through text that summarize descriptions, maps, and other images. The payment vehicle is a tax increase that ranges from $20 to $750 that would be paid annually for 5 y.

von Haefen et al. (10) value local water quality in North Carolina and vary the number of murky water days, health risk, and ecosystem conditions. The levels of these attributes are displayed as pie charts with three slices (e.g., high, medium, and low) that vary in size. The payment vehicle is an additional monthly fee that ranges from $4 to $32 that would be paid in perpetuity.

Lupi et al. (9) consider water quality improvements in Michigan and vary water quality as it supports wildlife, human contact (e.g., swimming) and recreational fishing. In addition, they test for differences with a two-attribute treatment where the human contact and recreational fishing attributes are combined into one attribute. The cost attribute is a one-time increase in the income tax that ranges from $45 to $965.

Johnston et al. (13) vary water quality in the Northeast with endpoints described as water pollution, aquatic life, and water safety (for human uses). The levels of these attributes are communicated with clickable maps that describe spatial scale and bar charts that display the magnitudes of the changes. The cost attribute is taxes and fees that range from $30 to $1,200 and would be paid annually in perpetuity.

Given the different definitions and measures of water quality and the wide range of cost amounts presented to survey respondents, it is clear that the four studies are attempting to value different goods and comparison of willingness to pay estimates across study might prove difficult. Further, these complicated and contrasting designs raise interesting statistical questions which are discussed below.

Sampling. Methods for surveying the general public have evolved considerably over since they first started to be widely deployed in the 1930s and now seem to be at a crossroads, if not a crisis. The long-standing gold standard of in-person interviewing utilizing a full probability sample have simultaneously become prohibitively expensive at the same time response rates have dramatically fallen, raising issues of representativeness. Telephone survey utilizing random digit dialing of landlines, long the mainstay of much applied work in marketing and political realms, now face severe problems from answering machines and the rapidly increasing fraction of the public only using a mobile phone. Mail surveys, which have often been seen as good for nonmarket valuation studies for their better ability (relative to telephone surveys) to present visual displays of the relevant policy changes, have long been known to suffer from potential interest-related sample selection issues since respondents get to glance through the questionnaire before answering it. To that problem, mail surveys, particularly to large general population samples, are also experiencing falling response rates. This has pushed researchers toward survey implementation over web-based platforms, which shares the same ability to display visual material that in-person and mail surveys have and to control the order in which question are seen that in-person surveys. Complex experimental designs, which are possible with in-person and mail surveys by printing different versions of a questionnaire, tend to be easier to implement in web-based surveys. Complicated skip patterns, long the forte of in-person surveys are, if anything are easier to reliably implement in web-based survey.

The long-standing problem with web-based surveys is, of course, the representativeness of the sample used. Early studies often lacked representation of the general public (19). The situation is now considerably different with three distinct approaches being used. The first used by several of the large firms dominant in the marketing and political spheres is to recruit an enormous number of people from a wide variety of sources to be in their panels who are incentivized to answer the firm's surveys online. Samples representative of the population of interest (e.g., American's 18 and over) are then drawn using probabilistic weighting schemes. Second, rather than recruit for a panel widely, it is possible to define the appropriate sampling frame, randomly choosing individuals from it and then heavily investing in trying to recruit members for a panel that can be repeatedly surveyed online for different purposes. A third approach is to recruit a group of respondents to participate in a special purpose survey online. The sampling element here also comes from the recruitment process, which is typically via selection of a random sample to mail solicitation to participate in the online survey from a master postal address list. It is often implemented in local areas where the effective sample size available from panels of the first two types is inadequate. A commonly used term for this approach when implemented using this recruitment mechanism is mail-push-to-web.

The comparative quality of the survey data collected via web-based surveys using these three different approaches is a topic that has been and is currently being intensively examined in the survey research literature resulting in improvements to all three approaches. The growing consensus is that the quality of a well-done web-based survey is on par with what can be achieved at this point in time with the other well-established survey modes.

Each of the studies in this Symposium use web-based survey platforms implemented using different approaches. Lupi et al. (9) use a sample drawn from Qualtrics' opt-in panel, one of the exemplars of the first approach. Vossler et al. (11) deploy the second approach by using the National Opinion Research Center (NORC) probability-based research panel, which is known to be of high quality with respect to the two metrics panels of this type are measured on, response rate to the initial full-probability based recruitment effort, and the propensity to respond to the survey they are sent. Two of the studies, von Haefen et al. (10) and Johnston et al. (13) use different address-based mail- push- to- web surveys.

All noncompulsory surveys are subject to bias, and internet surveys are no exception (20, 21). However, a common characteristic of mail push-to-web approaches is that only a relatively small proportion of those contacted actually complete the survey. In these cases, after accounting for undeliverable mailings and those responses that failed quality screening tests, the response for von Haefen et al. (10) was 15.2% and Johnston et al. (13) record a 5.1% response rate. Such rates would not be an issue if they were a random and hence representative selection of those approached. However, as von Haefen et al. (10) note, "Compared to county-level Census sociodemographic data, survey respondents were wealthier, whiter, more educated, and younger." As the authors highlight, such bias necessitates the use of weights to adjust the sample to better approximate the underlying population. However, alongside socioeconomic dimensions, surveys of goods which are spatially located are likely to have response rates which are biased toward those that live near to those sites (22). Those respondents are in turn likely to hold higher values than others in the population. A failure to adjust for the upward pressure these factors exert on sample means can lead to significant overestimation of aggregate values for a population (23). Vossler et al. (11) provide an innovative approach for adjusting for both sociodemographic and geographical biases through the joint use of a representative internet panel survey combined with small zone, distance weighted, value aggregation repeated across their overall study area. Combining this with information on variation in the baseline water quality yields spatially sensitive mapping of the values of water improvement policies which provide in-depth decision support for policy makers.

With low response rates using traditional modes of survey administration, web-based surveys are increasingly the norm as their quality has improved and properties have become better understood. The papers in this Symposium all make substantial contributions to understanding the implementation of nonmarket valuation surveys and the analysis of data from them. Their results suggest a rich research agenda on aspects of nonresponse and selection.

Econometric Approach. When the desire is to understand individual level behavior, the state-of-the-art with multiple response stated preference quasipanel data are to employ some type of empirical model that captures preference heterogeneity. The most common is some type of random parameter, aka mixed, logit models, that allow utility function preference parameters to follow assumed distribution with parameters (e.g., normal with unknown mean and variances) to be estimated. Perhaps the most common alternative is a latent class, aka finite mixture, logit model where instead of the typically assumed smooth distribution of preference parameters in the mixed logit model, there are a finite [small (2 to 5)] number of types or segments as they are referred to in the marketing literature. Two of the studies in the Symposium elicit one response per respondent and, while this format enhances incentive compatibility, are limited in the range of empirical models that can be employed. One of these single response studies employs a fixed coefficient logit [Lupi et al. (9)] while the other reports results from a Bayesian logit [Johnston et al. (13)]. von Haefen et al. (10) and Vossler et al. (11) report results from a mixed logit regression model that accounts for preference heterogeneity.

All four studies find that cost affects the choice negatively and water quality improvements affects the choice positively, as expected. Considering the two papers that use panel data and mixed logit models, the distribution of the coefficient estimates across the survey respondents could provide insights into how each definition of water quality is perceived by respondents. For example, a water quality attribute with a low ratio of the mean coefficient to the SD of the coefficient could suggest that there is much preference heterogeneity for that attribute or that attribute is perceived in various ways by the respondents. Information about the statistical distribution of parameters on water quality measures could be useful in determining the efficacy of the different approaches.

Testing the Validity of Findings

Caught between the need to ensure public goods are given due valuations in decision-making and the lack of a criterion yardstick, economists are forced to rely on a variety of less definitive tests. While much attention is paid to whether estimated values converge on those for similar goods as reported elsewhere in the academic literature, such examinations are ultimately unsatisfying because they mitigate against new and improved studies which might provide better, yet different, values to those in the prior literature. Indeed if the prior literature were infallible then this would negate the need for further valuations. Therefore, a more defensible approach is to examine whether estimated values vary according to prior expectations. One source of such expectations concerns the responsiveness of values to variation in the characteristics of individuals themselves. For example, do those with higher incomes and therefore less constraint on their ability to pay generally exhibit higher WTP values? Do those who rarely visit beaches have lower WTP than those who frequently use them for recreation?

Tests of the variation of WTP with individual characteristics are routinely satisfied. However, the papers collected in this Symposium conduct further analyses of the responsiveness of values to changes in the physical definition of the environmental public goods under assessment. Again we can test against expectations, although now these are often more nuanced. For example, we might expect that as the scale or "scope" of the good on offer increases in quantity and/or quality so valuations might rise. This expectation might well hold over a range of changes; however, economic theory shows that individuals will at some point become satiated by the level of a good offered to them. So, for example, if all lakes are polluted then an individual might have a high WTP for cleaning up a first lake. But once a first pristine lake has been provided, that individual is likely to have a lower WTP to clean up a similar second lake. This phenomena of a per unit (or "marginal") WTP which declines with successive units of provision gives us a test of the validity of stated preference responses although it does mean that such a scope test (16–18) does not have the simple expectation of constant unit values and that eventually marginal WTP may diminish to zero. A related second expectation, which provides a further test of validity, is that individuals may well care more about changes that occur near to them than far away (indeed this is the essence of the revealed preference travel cost method; holding site quality and other factors equal then visitors would prefer to undertake lower cost, nearby visits). So, continuing the lake example, individuals may hold higher values for improvements to lakes near to their home (which they are more likely to visit and recreate at) than those in distant locations. The expected "distance decay" in such use values provides a further validity test for assessing stated preference responses (22, 24, 25). One might expect that nonuse values would be immune to such distance effects; if I care about wildlife that I never see, why should it matter if it is located near to me or on the other side of the world? However, empirical results find that distance decay can also be observed in nonuse values, albeit often to a weaker extent than for use values (23, 26). This would be expected to occur, for instance, if respondents feel more of an affinity for protecting local biodiversity or that others in more distant lands should be looking after their wildlife. Whatever the reason, again this can provide a further test of validity.

Relatedly, individuals may also have higher WTP for water quality at sites they visit than at those they do not visit. This is likely correlated with distance, but the revealed preference literature has established that the highest-valued sites are not necessarily those closest to individuals' homes and that accounting for visitation patterns makes a large difference in the magnitude of hedonic property estimates of the WTP for water quality (27). This provides an additional test for validity of use-value estimates in the stated preference literature.

The papers presented in this Symposium represent a new high point in assessments of the validity of stated preference responses. In particular, the examination against expectations regarding the impact of variation in the scope and spatial distribution of changes in ecosystem services upon derived WTP values is exemplary. As might be expected from its title, the paper by Johnston et al. (13), "Spatial Dimensions of Water Quality Value in New England River Networks," makes such validity testing a central feature of its analysis. Their use of interactive GIS maps builds on prior choice experiments which show that visual presentations may help tease out distance decay effects by enhancing respondents' abilities to comprehend information which might be challenging if presented through a purely numeric format (28, 29). Results confirm expectations regarding both distance decay in values and their sensitivity to the quality of improvements. However, while von Haefen et al. (10) report value sensitivity to changes in quality and lower values for more distant respondents, the latter effect was not statistically significant. This may be due to the fact respondents were only asked questions about their own county, so the degree of spatial variation is small. This contrast is most marked in comparison to the Vossler et al. (11) study which considered changes occurring both in local subcatchments and in watersheds across multiple states. This larger canvas allows the authors to estimate values for potential policies that vary in their location, spatial scope, and the extent of the water quality change. Results find that all of these dimensions exert expected effects upon valuation results. Clear evidence of scope sensitivity with eventual satiation is reported with respondents initially increasing their WTP with the size of watershed improved but marginal WTP falling toward zero as successively larger improvements are offered. With regard to distance decay Vossler et al. (11) find that people are willing to pay twice as much for an improvement policy that targets their home watershed than one affecting a more distant watershed. Nevertheless, analysis of payments for such distant improvements provides considerable evidence of nonuse values being an element of the total value of water quality improvements.

As noted above, WTP may also vary with the degree to which respondents visit or otherwise value the sites in question, which may be correlated with distance, but not necessarily perfectly so. Johnston et al. (13) demonstrate that people value water quality at specific locations, only some of which may be near their own homes. In concert with recent evidence from the revealed preference literature (27), their work suggests that failing to identify the specific water bodies that people value [for whatever reason(s)] may impair researchers' ability to detect the signal of demand for water quality at such locations amidst the noise from a multitude of monitoring locations on water bodies where such demand is low or nonexistent. It would be instructive to explore whether the valued locations that Johnston et al. (13) identify through their creative use of technology to track respondents' attention correlate with observable patterns of recreational visitation or other potential drivers of willingness to pay. This paper also forges a fascinating hybrid between revealed and stated preference, as respondents "reveal" through their interaction with online maps—visitation with their eyes instead of with their feet in the manner exploited for many decades by researchers estimating the recreational demand models originally envisioned by Hotelling (30). Intriguingly, Vossler et al. (11) also differentiate survey respondents' valuation of water quality improvements in their home watershed vs. elsewhere, finding that WTP for water quality outside of a respondent's home watershed is much smaller than that for local improvements. A small extent of the market for water quality is somewhat inconsistent with the recreation demand literature (31). Future research should explore whether the findings from these two papers can be reconciled using an approach that adopts the BCG Ladder piloted by Vossler et al. (11) but uses recreational visitation patterns or other behaviors to identify the particular waters that respondents may value.

Overall Contributions to Public Policy

This Symposium combines two important aspects of environmental benefit valuation—one methodological and one topical. The papers focus on the intersection of the valuation of changes in ambient water quality, and "stated preference" methods, which use carefully structured surveys and related experimental techniques to elicit estimates of individuals' demand for environmental amenities (or the damages from disamenities). The papers' illumination of issues at this intersection is timely in an applied sense, given seismic shifts in the jurisdictional scope of the Clean Water Act in the United States during the current and last two presidential administrations (32), and an upcoming decision from the US Supreme Court in Sackett v. EPA which is likely to continue the quaking. It is also of great academic interest, given major advances in stated preference methods in recent decades (33), and the new ground that the Symposium papers' authors break in their development.

Role of Stated Preference Approaches to Valuing Water Quality. Stated preference approaches have been used for water quality valuation in regulatory processes in various ways over time. For some water quality rules, including some of the earliest applications of regulatory benefit-cost analysis to water pollution rules, the EPA has used contingent valuation estimates from the national study by Mitchell and Carson (34) and a small number of other location-specific studies to account for recreation and nonuse values (1). More recently, the EPA has used more sophisticated practices for this kind of "benefit transfer," estimating nonuse benefits via meta-analysis (35). In this process, the EPA uses a benefit function to extrapolate from a large set of estimates in the literature to the specific policy case. A recent example is the use of a peer-reviewed meta-analysis of stated preference estimates of the value of wetland acreage (36), commissioned by the EPA for its economic analysis of rules attempting to clarify the Clean Water Act (CWA) jurisdiction (the boundaries of the "Waters of the United States").

Stated preference methods play a key role in water quality valuation for several reasons. First, many individuals value water quality even in waters they never expect to encounter for recreation or other purposes. Consider, for example, the damages to water quality in the Gulf of Mexico from the 2010 Deepwater Horizon/Macondo disaster and those to Prince William Sound off the coast of Alaska from the 1989 *Exxon Valdez* oil spill. Careful research demonstrates that individuals across the United States experienced a loss when these events occurred, even if they did not incur direct damages related to recreational, commercial, or other uses of the affected waters (37, 38). As the authors of several of the Symposium papers point out, such "nonuse" or "passive use" values do not leave footprints in markets, thus stated preference methods are the only way to estimate them. For water quality policies with significant nonuse benefits, stated preference methods are therefore essential valuation techniques.

Second, for benefit–cost analysis of regulation, natural resource damage assessment in the context of litigation, and in other settings, policymakers and other stakeholders may seek a comprehensive estimate of the value of water quality. The alternatives to stated preference techniques—"revealed preference" methods which exploit observational data on water quality and specific endpoints it affects, such as property prices, recreational visitation patterns, and human health, to estimate willingness to pay for water quality—are generally designed to capture one or a small number of types of values. In contrast, a single stated preference survey can, in principle, capture as many aspects of water's value as the researcher designs it to capture (including nonuse values). While the two approaches are complementary, their capacity for comprehensiveness make stated preference approaches an important part of the water quality valuation toolkit.

Third, one challenge in estimating the value of water quality is understanding what the affected individuals know about the implications of changes in a waterbody's appearance, smell, and other noticeable characteristics, and how those changes relate to water quality parameters that can be easily observed by the researcher at a broad spatial and intertemporal scale. Unlike revealed preference methods, stated preference approaches involve an interaction between the researcher and the respondent via the survey questions being asked. This provides a unique opportunity to explain the change in the good being valued and the implications of that change. This interaction makes willingness to pay sensitive to descriptions of the good, one among many sources of variation, which can be addressed through careful research design (33). However, it is also an advantage of the approach relative to methods that estimate value based on impacts to property values or recreation behavior in a context where the affected individuals lack perfect knowledge of both water quality and its implications for their health and other aspects of well-being.

Confronted with these advantages, it is not surprising that the most comprehensive assessment of the benefits of the CWA in the United States to date (39) used contingent valuation, a stated preference approach. Given major methodological advances over the ensuing three decades, changes in baseline water quality and the spatial distribution of water pollution challenges, and potential differences in which aspects of water quality are valued by individuals and firms, where, and how much, this Symposium lays important groundwork for much-needed new water quality valuation work at broad spatial scales.

Contributions of the Symposium Papers. In combination, these Symposium papers expand economists' toolkit for estimating the benefits of water quality improvements, or conversely the damages from water quality degradation, in multiple directions. They offer improved water quality measures and stronger links between what can be measured directly in the environment and what is valued (and where) by individuals. They provide avenues to better capture nonuse values associated with water quality, which may be substantial in many settings. They tap new data sources and new technologies that help define the spatial extent of the market for water quality. These tools and other contributions will spur exciting new basic research into the value of water quality and help support better and more comprehensive valuation of water pollution control policies in regulatory and other applied settings.

Data, Materials, and Software Availability. There are no data underlying this work.

1. C. Griffiths *et al.*, Environmental Protection Agency valuation of surface water quality improvements. *Rev. Environ. Econ. Policy* **6**, 130–146 (2012).
2. A. Sinden, The problem of unquantified benefits. *Environ. Law* **49**, 73–129 (2019).
3. D. R. Petrolia *et al.*, Nonmarket valuation in the Environmental Protection Agency's regulatory process. *Appl. Econ. Perspect. Policy* **43**, 952–969 (2021).
4. C. L. Kling, R. W. Arritt, G. Calhoun, D. A. Keiser, Integrated assessment models of the food, energy, and water nexus: A review and an outline of research needs. *Annu. Rev. Resour. Econ.* **9**, 143–163 (2017).
5. C. C. Carey *et al.*, Advancing lake and reservoir water quality management with near-term, iterative ecological forecasting. *Inland Waters.* **12**, 107–120 (2022), 10.1080/20442041.2020.1816421.
6. B. L. Keeler *et al.*, Linking water quality and human well-being for improved assessment and valuation of ecosystem services. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 18619–18624 (2012).
7. N. I. McClelland, "Water quality index application in the Kansas River Basin" (Region VII, EPA-907/9-74-001, US Environmental Protection Agency, 1974).
8. R. C. Mitchell, R. T. Carson, "An experiment in determining willingness to pay for national water quality improvements" (US Environmental Protection Agency, Resources for the Future, Washington, DC, 1981).
9. F. Lupi, J. A. Herriges, H. Kim, R. J. Stevenson, Getting off the ladder: Disentangling water quality indices to enhance the valuation of divergent ecosystem services. *Proc. Natl. Acad. Sci. U.S.A.* (2023) (in submission).
10. R. H. G. von Haefen *et al.*, Estimating the benefits of stream water quality improvements in urbanizing watersheds: An ecological production function approach. *Proc. Natl. Acad. Sci. U.S.A.* (2023) (in submission).
11. C. A. Vossler *et al.*, Valuing improvements in ecological integrity in local and regional watersheds: The biological condition gradient ladder. *Proc. Natl. Acad. Sci. U.S.A.* (2023) (in submission).
12. R. A. Hill *et al.*, Estimating biotic integrity to capture existence value of freshwater ecosystems. *Proc. Natl. Acad. Sci. U.S.A.* (2023) (in submission).
13. R. J. Johnston *et al.*, Spatial dimensions of water quality value in New England river networks. *Proc. Natl. Acad. Sci. U.S.A.* (2023) (in submission).
14. R. T. Carson, T. Groves, Incentive and informational properties of preference questions. *Environ. Resour. Econ.* **37**, 181–210 (2007).
15. R. T. Carson, T. Groves, "Incentive and information properties of preference questions: Commentary and extensions" in *The International Handbook on Non-market Environmental Valuation* (Edward Elgar Publishing, Cheltenham, 2011).
16. V. K. Smith, L. L. Osborne, Do contingent valuation estimates pass a "scope" test? A meta-analysis. *J. Environ. Econ. Manage.* **31**, 287–301 (1996).
17. J. C. Whitehead, Plausible responsiveness to scope in contingent valuation. *Ecol. Econ.* **128**, 17–22 (2016).
18. N. Borzykowski, A. Baranzini, D. Maradan, Scope effects in contingent valuation: Does the assumed statistical distribution of WTP matter? *Ecol. Econ.* **144**, 319–329 (2018).
19. H. Lindhjem, S. Navrud, Using internet in stated preference surveys: A review and comparison of survey modes. *Int. Rev. Environ. Resour. Econ.* **5**, 309–351 (2011).
20. M. van Selm, N. W. Jankowski, Conducting online surveys. *Qual. Quan.* **40**, 435–456 (2006), 10.1007/s11135-005-8081-8.
21. Y. Khazaal *et al.*, Does self-selection affect samples' representativeness in online surveys? An investigation in online video game research. *J. Med. Int. Res.* **16**, e164 (2014), 10.2196/jmir.2759.
22. I. J. Bateman, I. H. Langford, Non-users' willingness to pay for a National Park: An application and critique of the contingent valuation method. *Reg. Stud.* **31**, 571–582 (1997), 10.1080/00343409750131703.
23. I. J. Bateman, B. H. Day, S. Georgiou, I. Lake, The aggregation of environmental benefit values: Welfare measures, distance decay and total WTP. *Ecol. Econ.* **60**, 450–460 (2006).
24. I. J. Bateman, I. H. Langford, N. Nishikawa, I. Lake, The Axford debate revisited: A case study illustrating different approaches to the aggregation of benefits data. *J. Environ. Plann. Manage.* **43**, 291–302 (2000).
25. B. H. Day, I. J. Bateman, A. Binner, S. Ferrini, C. Fezzi, Structurally-consistent estimation of use and nonuse values for landscape-wide environmental change. *J. Environ. Econ. Manage.* **98**, 102256 (2019).
26. N. Hanley, F. Schläpfer, J. Spurgeon, Aggregating the benefits of environmental improvements: Distance-decay functions for use and non-use values. *J. Environ. Manage.* **68**, 297–304 (2003).
27. Y. Kuwayama, S. Olmstead, J. Zheng, A more comprehensive estimate of the value of water quality. *J. Public Econ.* **207**, 104600 (2022).
28. I. J. Bateman, B. H. Day, A. P. Jones, S. Jude, Reducing gains/loss asymmetry: A virtual reality choice experiment (VRCE) valuing land use change. *J. Environ. Econ. Manage.* **58**, 106–118 (2009), 10.1016/j.jeem.2008.05.003.
29. T. Badura, S. Ferrini, M. Burton, A. Binner, I. J. Bateman, A new approach to capturing the spatial dimensions of value within choice experiments. *Environ. Resour. Econ.* **75**, 297–322 (2020), 10.1007/s10640-019-00358-3.
30. H. Hotelling, "Letter (18 June 1947) quoted by Prewitt, R.E." in *The Economics of Public Recreation: An Economic Study of the Monetary Evaluation of Recreation in National Parks* (United States National Park Service, Land and Recreational Planning Division, Washington, DC, 1949).
31. F. Lupi, D. J. Phaneuf, R. H. von Haefen, Best practices for implementing recreation demand models. *Rev. Environ. Econ. Policy* **14**, 302–323 (2020), 10.1093/reep/reaa007.
32. D. Keiser *et al.*, A water rule that turns a blind eye to transboundary pollution. *Science* **372**, 241–243 (2021).
33. R. J. Johnston *et al.*, Contemporary guidance for stated preference studies. *J. Assoc. Environ. Resour. Econ.* **4**, 319–405 (2017).
34. R. C. Mitchell, R. T. Carson, "*Willingness to pay for national freshwater quality improvements*" (CR 810224–01–4, Draft Final Report to the US EPA, 1984).
35. W. J. Wheeler, "Benefit transfer for water quality regulatory rulemaking in the United States" in *Benefit Transfer of Environmental and Resource Values: The Economics of Non-Market Goods and Resources*, R. J. Johnston, J. Rolfe, R. S. Rosenberger, R. Brouwer, Eds. (Springer Science, Berlin, 2015).
36. K. Moeltner, J. Balukas, E. Besedin, B. Holland, Waters of the United States: Upgrading wetland valuation via benefit transfer. *Ecol. Econ.* **164**, 106336 (2019).
37. R. T. Carson *et al.*, Contingent valuation and lost passive use: Damages from the Exxon Valdez oil spill. *Environ. Resour. Econ.* **25**, 257–286 (2003).
38. R. C. Bishop *et al.*, Putting a value on injuries to natural assets: The BP oil spill. *Science* **356**, 253–254 (2017).
39. R. C. Mitchell, R. T. Carson, *Using Surveys to Value Public Goods: The Contingent Valuation Method* (Resources for the Future Press, Washington, DC, 1989).