Measuring Benefits from Spatially-Explicit Surface Water Quality Improvements:

The Roles of Distance, Scope, Scale, and Size

Dong Soon Choi[1], Richard Ready[2,3]

[1] Department of Agricultural Economics, Sociology and Education, Pennsylvania State University

[2] Department of Agricultural Economics and Economics and Montana Institute on Ecosystems, Montana State University

[3] Corresponding Author:
309E Linfield Hall
Montana State University
Bozeman, MT 59717
Richard.ready@montana.edu
406-994-7365

Highlights

- In our preferred sample, we find that the value to a household from an improvement in water quality declines roughly linearly with distance, and reaches zero at between 30 and 40 miles.
- We find no evidence that willingness to pay for a water quality improvement is nonlinear in the scale or scope of the improvement.
- We find that independent valuation and summation of small, local programs to improve water quality is empirically justified.

Keywords: water quality, nonmarket valuation, willingness to pay, distance decay, scope test

JEL Codes: Q5

This research was funded by a grant from the Pennsylvania Department of Environmental Protection. The views expressed here are solely those of the authors, and do not necessarily represent the policy or position of the Department of Environmental Protection.

Measuring Benefits from Spatially-Explicit Surface Water Quality Improvements:

The Roles of Distance, Scope, Scale, and Size

Abstract

Programs to improve water quality do not improve all water bodies equally. Evaluation of the benefits of such programs must account for spatial heterogeneity in the results. This study uses a choice experiment survey to explore how the value to a household of a surface water quality improvement varies as a function of (i) the distance between the household and the affected streams and rivers, (ii) the degree to which the quality of the water has been improved, (iii) how many stream and river miles have been improved, and (iv) the size of the affected streams or rivers. Results show evidence that value declines with distance in an approximately linear way, weak evidence that large rivers are worth more than small rivers, and no evidence that willingness-to-pay is nonlinear in either the degree of water quality improvement or the number of stream miles improved. These results indicate that it may be defensible in applied work to value small, spatially-explicit water quality improvement projects independently and then sum over projects.

Measuring Benefits from Spatially-Explicit Surface Water Quality Improvements:

The Roles of Distance, Scope, Scale, and Size

1. Introduction

Cost-benefit analyses of policies that aim to improve surface water quality need to account for the fact that water quality improvements happen in specific places. Local actions, such as remediation of a specific watershed degraded by industrial or mining activity or adoption of agricultural best management practices to reduce sediment and nutrient runoff into streams, result in specific water quality improvements in specific stream and river reaches. But even regional and national policies, such as regulations promulgated under the Clean Water Act, will affect different streams and rivers differently, with some experiencing larger improvements and others smaller improvements.

Because water quality improvements do not occur uniformly over the landscape, the values generated by those improvements will not be uniform over all households. Unfortunately, much of the research that has been conducted valuing surface water quality improvements has valued uniform, region-wide improvements on the scale of a river basin (Loomis et al. 2000; Hanley et al. 2006), a state (Hite et al. 2002), or the entire country (Carson and Mitchell 1993). In order to estimate the benefits to a specific household from a spatially-explicit improvement in water quality, it is necessary to know how value varies with the types of the specific water bodies improved, the degree to which water quality has been improved in those water bodies, and how the spatial relationship between the household and the improved water bodies.

This study asks questions in four research areas relevant to this issue.

(i) Distance – How does the value of a water quality improvement vary as a function of the distance between the household and the affected streams and rivers? Is this relationship linear or nonlinear? Is there a threshold distance beyond which households receive no value?

(ii) Scope – How does the value of a water quality improvement vary as a function of the degree to which the quality of the water has been improved? Is a larger reduction in nutrient and sediment concentrations worth more than a smaller reduction? Is this relationship proportional?

(iii) Scale – How does the value of a water quality improvement vary as a function of how many stream and river miles have been improved? Is an improvement to 20 miles of streams worth more than an improvement to 10 miles? Is this relationship proportional?

(iv) Size – How does the value of a water quality improvement vary as a function of the size of the affected streams or rivers. Programs to improve water quality often have larger impact on smaller streams than on larger rivers. Are improvements to smaller streams worth less than improvements to larger rivers?

These questions are explored using a stated choice valuation survey of water quality improvements in three watersheds in Pennsylvania. The answers to these questions have important implications for how cost-benefit analyses should be conducted both of programs aimed at improving water quality in specific locations as well as for regulatory impact analyses of nation-wide regulations.

2. Literature Review

As a point of departure in the following literature review, recent U.S. Environmental Protection Agency (EPA) practices when conducting regulatory impact analyses (RIA) of major rules under the Clean Water Act will be highlighted. Two specific recent RIAs will be discussed, an RIA conducted in 2009 (EPA 2009a, 2009b) for effluent guidelines and standards for the construction and development category (referred to here as the C&D RIA) and an RIA conducted in 2015 (EPA 2015) for effluent guidelines and standards for the steam electric power generating point source category (referred to here as the Steam Electric RIA).

2.1. Measuring Water Quality

Water quality includes multiple dimensions, with a variety of quality indicators that vary regionally due to differences in data availability, diverse local conditions, and differing methodological assumptions (Griffiths et al. 2012). Several attempts have been made to construct an objective water quality index (WQI) that summarizes information on multiple quality indicators. Notable efforts include the Oregon Water Quality Index, the Canadian Council of Ministers of the Environment Water Quality Index, and National Sanitation Foundation Water Quality Index. While there is no clear consensus on a single WQI, the literature has tended to favor use of the National Sanitation Foundation WQI (Griffiths et al. 2012), and that is the WQI used by the EPA in both their 2009 C&D RIA and their 2015 Steam Electric RIA.

The National Sanitation Foundation WQI is a continuous index of water quality that ranges from 0 to 100 (where 0 is poor and 100 is excellent) based on up to nine water quality parameters. For a given stream segment, each water quality parameter is measured or modeled, and the stream is rated according to its percentile among all streams for that parameter, with the best stream given a rating of 100 and the worst stream a rating of 0. The WQI then aggregates these indicator-specific ratings as follows,

$$WQI = \sum_{i=1}^{9} (q_i)^{w_i}$$

where q_i is the stream segment's percentile for parameter i and w_i is the weight assigned to parameter i. Weights are assigned based on expert judgement, and sum to one. The WQI can be applied with fewer than the original nine quality parameters in situations where information on all nine is limited, with the appropriate re-scaling of the weights (Tyagi et al. 2013). The WQI is widely used, particularly in the United States (Lumb et al. 2011) and it is the most commonly used compound indicator by the EPA (Griffiths et al. 2012). Van Houtven et al. (2007) also advocate the use of the WQI because of its flexibility at the national level and its compatibility with benefit transfer.

A difficulty of the WQI for stated preference valuation purposes is that survey respondents may not find its numerical values meaningful. Mitchell and Carson (1989) and Smith and Desvouges (1986) were the first to use a Water Quality Ladder (WQL) as an aid to respondents in their stated preference surveys (Walsh and Wheeler 2013). The WQL presents respondents with a series of water quality levels based on the supported uses: drinkable, swimmable, fishable, and boatable. Steps of the WQL are typically assigned specific WQI values, in order to allow translation from water quality model output to the WQL. Using a 100-point WQI scale, it is common practice to use WQI values of 90 or 95 as the threshold for the drinkable condition, 70 for swimmable, 50 for fishable, and 25 for boatable, based on Vaughan (1981).

While the WQL is useful for communicating water quality changes to respondents who lack a technical background, it is not appropriate for valuing specific changes in water quality that might result from specific policies, because it places high values on small changes that occur at the thresholds between categories while failing to capture the benefits of changes that occur from improvements within categories (EPA 2002). Instead, the EPA in its RIAs has used a two-step process to value changes in water quality that projected to occur as a result of new regulations. First, it conducts a meta-analysis of water quality valuation studies. In studies that valued discrete changes along a WQL, the water quality changes valued are converted to numerical WQI scores. The meta-analysis generates a WTP value function calibrated to the WQI. It then uses that value function to value the WQI changes predicted from the proposed rule.

2.2 Independent Valuation and Summation

If values for surface water quality improvements are nonlinear in the scope and/or scale of the improvements, for example if willingness to pay exhibits decreasing marginal value in either scope or scale, then individual interventions cannot be valued independently and then summed together. Instead, each intervention would have to be evaluated taking into account all previous interventions, and sequencing would have an important impact on the value of any specific intervention. This issue is a well-recognized theoretical concern (Hoehn and Randall 1989; Bateman et al. 1997; Boyle et al. 1994; Eftec 2009; Hanley et al. 2003; Kling and Phaneuf 2018).

If we value two independent programs, A and B, that each improve water quality, it is generally not appropriate to independently value the two and then sum their values. That is, WTP for program A plus WTP for program B will generally be larger than WTP for a combined program A+B. This is for two reasons. First, there are potential income effects. If a household is required to pay their WTP for program

A, they will have less money left over to pay for program B. While this is a valid theoretical concern, typical values for water quality improvements considered here are small relative to total income, so income effects are anticipated to be small. The second reason why independent valuation and summation can overstate total WTP for larger programs is due to the potential for decreasing marginal value. A household may place a high value on improving some nearby streams, in order to have some opportunity for recreation. However, having obtained some improved streams, they may value additional improvements less highly. Similarly, a household may value an improvement in a stream that makes it suitable for various uses, but may place lower value on subsequent improvements once those uses are supported.

Substitution effects and income effects would show up as nonlinearities between the value of a water quality improvement and either the scale and or the scope of the improvement. If nonlinearities exist in scale or scope, then it is not defensible to value projects and programs independently.

2.3. Linearity in scope

Evidence whether the value of an improvement in water quality is proportional[1] to the scope of the improvement (i.e. the size of the increase in the WQI) is mixed. Van Houtven et al. (2007) conducted a meta-analysis of available water quality valuation studies and found that a larger improvement in water quality is worth more. However, they found some evidence in some of their models of decreasing marginal value as the scope of the WQI improvement increases, but only for very large improvements in water quality (increases in the WQI of 60 points). EPA's 2015 meta-analysis (EPA 2015) finds that WTP per point increase in the WQI is smaller when the valued change in the WQI is larger. These findings suggest that, at least for large changes in the WQI, the relationship between value and increases in the WQI is positively sloped but concave, i.e. that marginal WTP for WQI decreases as the change in the

[1] A proportional relationship is one where value is a linear function of the scope or scale of the improvement, with zero intercept. Some meta-analyses use a value function that does not impose a zero intercept, implying that a 0-point improvement in WQI or a program that affects 0 water bodies could have positive value. We use the term "nonlinear" to include value functions that have curvature as well as value functions that have non-zero intercept.

WQI gets larger. The meta-analysis conducted for EPA's 2009 C&D RIA (EPA 2009a) assumed a structural form that imposes decreasing marginal value for larger improvements.

A related issue is whether value for an improvement in the WQI depends on the baseline value of the WQI. Here, results are mixed. Ge et al. (2013), in a meta-analysis, found that people are willing to pay more for an improvement in waterbodies with bad initial conditions than those with already good initial conditions, suggesting diminishing marginal value with respect to the change in water quality. In contrast, Houtven et al (2007) find no evidence that willingness to pay for an improvement in water quality varies with the baseline level of water quality. EPA's 2015 meta-analysis found that, when the size of the WQI change is controlled for, WTP does not depend on the baseline WQI. Curiously, Georgiou et al. (2000), in a study measuring the benefit of water quality improvements in the River Tame, UK, finds that WTP increases at an increasing rate with improvements changes in WQI, suggesting increasing marginal value. For a given water body, it could be that there are value thresholds, and that marginal value could be increasing over some ranges and decreasing over others.

Previous studies that explored linearity in the scope of the WQI improvement have tended to value large increases in the WQI. For example, the WQI increases valued in the studies used in EPA's 2015 meta-analysis ranged up to 50 points. Increases of that scope are uncommon in applied policy contexts. Projected WQI improvements from the 2009 C&D rulemaking were less than 0.5 WQI points for over 99% of stream reaches. Similarly, projected improvements from the 2015 Steam Electric rulemaking were less than 1 WQI point for over 99% of stream reaches. Targeted implementation of agricultural best management practices can result in somewhat larger improvements, but WQI improvements in the range of 10 to 50 points would only be expected in unique situations, for example where a severely degraded stream is remediated. In this study, we explore the issue of linearity in scope for a range of WQI improvements (up to 6 WQI points) that more closely matches improvements that might result from regional or national efforts to improve water quality.

In their 2009 C&D RIA, the EPA used a value function that was nonlinear in the scope of the change in the WQI. However, in their 2015 Steam Electric RIA, they used a linear value function, arguing that water quality changes were small enough that a linear approximation was valid.

2.4. Linearity in scale

Investigations of whether the value of an improvement in water quality is linear in the scale of the improvement (number of stream miles improved) is generally done by comparing WTP for a program with a smaller scale to WTP for a program with a larger scale. Such comparisons often find that value does not increase proportionally with the scale of the improvement. For example, Eftec (2010) found that a large program that improved twice as many stream miles was worth more than the smaller program, but not worth double a smaller program.

Meta-analyses have shown positive scale effects as well. EPA's 2009 meta-analysis found that WTP increased with the number of rivers and ponds valued in a study. Ge et al. (2013) find that WTP for improvements that affect a larger region is higher than WTP for improvements affecting a smaller region. Van Houtven et al (2007) found that WTP for improvements that only affect local water bodies is less than WTP for improvements that affect larger regions, but the estimated coefficient was not statistically significant. None of these meta-analyses tested whether the observed scale effect was proportional or nonlinear.

Such comparisons are complicated by the confounding role that distance plays when considering changes in scale. As the size of the region being improved increases, for example from a county to a state, it will include water bodies that are located farther from any given household. If improvements located farther away are worth less than improvements located closer, then it would be expected that value would increase nonlinearly with the scale of the improvement. The current study is one of few that varies the scale of the improvement in water quality while holding the distance between the improved water bodies and the household constant.

In both their 2009 C&D RIA and their 2015 Steam Electric RIA, the EPA treated individual streams independently and assumed that value was proportional to the number of stream miles improved within the state (in the 2009 C&D RIA) or within 100 miles of the household (in the 2015 Steam Electric RIA).

2.5. Stream size and value

The size (width, depth or flow) of the stream(s) river(s) being improved is rarely considered in the surface water quality valuation literature. Meta-analyses (Houtven et al 2007, EPA 2009a, EPA 2015), do not typically test whether value is related to stream or river size. Such a test would be difficult to implement, because many source studies value regional changes in water quality that affect large and small streams the same. In the context of lakes, there is some evidence that large lakes tend to generate more recreational use and hence more benefits to users than smaller ones (Ge et al 2013). But to our knowledge, similar evidence does not exist for streams and rivers. In their 2009 C&D RIA and their 2015 Steam Electric RIA, the EPA values all rivers and streams equally.

2.6. Distance decay in value

Two approaches are typically used to account for the spatial relationship between an improvement in environmental quality and the households who value that improvement. One approach defines a market area for the environmental good, i.e. the area within which households are assumed to hold value for the good. This might be a political jurisidiction (watershed, county, state) or the area within a specified distance cutoff from the improved water body or bodies. This approach is similar to how market area is defined for revealed preference studies, where only users living within a specific geographic region or within a specific distance of the resource (for example 200 miles) are included in the analysis (Parsons 2017). With this approach, all households within the market area are assumed to value the good in the same way, and households outside the market area are assumed to hold zero value for the good. Using this approach, aggregate values are often sensitive to the definition of the market area

(Loomis 1996, Vajjhala et al 2008). This is the approach taken both in EPA's 2009 C&D RIA, where they defined the market area as all households located within the same state as the improved water body, and in their 2015 Steam Power RIA, where they defined the market area as all households located within 100 miles of the improved water body. Within those market areas, households are assumed to value all streams and rivers equally.

In the second approach, a continuous distance decay relationship is estimated between value per household and distance from an improved water body. Several studies have explored how WTP for a water quality improvement declines with distance to the water body, but the results have been inconsistent. In particular, the shape of a distance decay function may vary depending on the context. Hanley et al. (2003) find that value for water quality improvements in a river in southern England decays with distance, but nonlinearly with a convex function. Further, they find that values held by households located near the river are dominated by use values while values held by households located more than 12 kilometers from the river are dominated by non-use values. They suggest using a 100 kilometer cutoff for aggregation of values. Bateman et al. (2000), in a study valuing preservation of the Norfolk Broads from the threat of saline flooding, impose a nonlinear distance decay function that approaches zero WTP asymptotically with increasing distance. However, they only aggregate values by zones out to a distance of 110 kilometer. Kim et al (2015), in a study valuing restoration of an urban stream in South Korea, also impose a nonlinear distance decay function. They found that value decreases with distance and was not significantly different from zero at a distance of 25 kilometers from the improved water body.

In contrast, Farber et al. (2000) and Georgiou et al. (2000) use linear distance decay functions. Georgiou et al. value inner city river water quality improvements for the River Tame in Birmingham, England, and find cutoff distances in the range of 20–28 kilometers, depending on the scope of the improvement. Farber et al. (2000) value water quality improvements in two sub-basins of the Lower Allegheny watershed in western Pennsylvania. Their results suggest that benefits decline to zero at a distance of between 42 and 54 miles (68 to 87 kilometers) from the improved streams.

Not all studies find that values decay with distance and/or that there exists a distance at which values fall to zero. Johnston and Ramachandran (2014) did not find statistically significant relationships between marginal values for most attributes and distance, though they did find spatial heterogeneity in values with values higher in specific "hot spots." Loomis (1996) found significant spatial decay in values for a dam removal project in Oregon, but found positive and economically meaningful values at distances greater than 1000 miles from the project.

3. Methods

To answer the research questions posed in the Introduction, we conducted a stated preference survey to measure WTP for specific water quality improvements located in specific places. A stated preference survey was used, as opposed to a revealed preference method such as the travel cost method or the hedonic pricing method, because stated preference methods are the only methods that can capture nonuse values (Alpizar et al. 2001). Also, a stated preference approach allows variation in scenarios that is not available in the real world. The choice experiment (CE) stated preference method was chosen over the contingent valuation method because it allows identification of marginal values along multiple attribute dimensions (Hanley et al. 2001). Further, the CE method has been found to show stronger sensitivity to scope than CV (Foster and Mourato 2003; Goldberg et al. 2007) and may minimize some of the response difficulties found in CV such as protest bids, strategic behavior, and yeah saying (Hanley et al. 2001).

3.1. Study Area

We identified three target watersheds all located within Pennsylvania and within the Susquehanna River basin, which drains into the Chesapeake Bay – the Mahantango Creek watershed, the Spring Creek watershed, and the Conewago Creek (east branch) watershed (Figure 1). These three watersheds were chosen as representative of the kind of watersheds where improvements are likely to be needed in order to reduce nitrogen, phosphorous and sediment loadings to achieve water quality goals in the Chesapeake Bay.

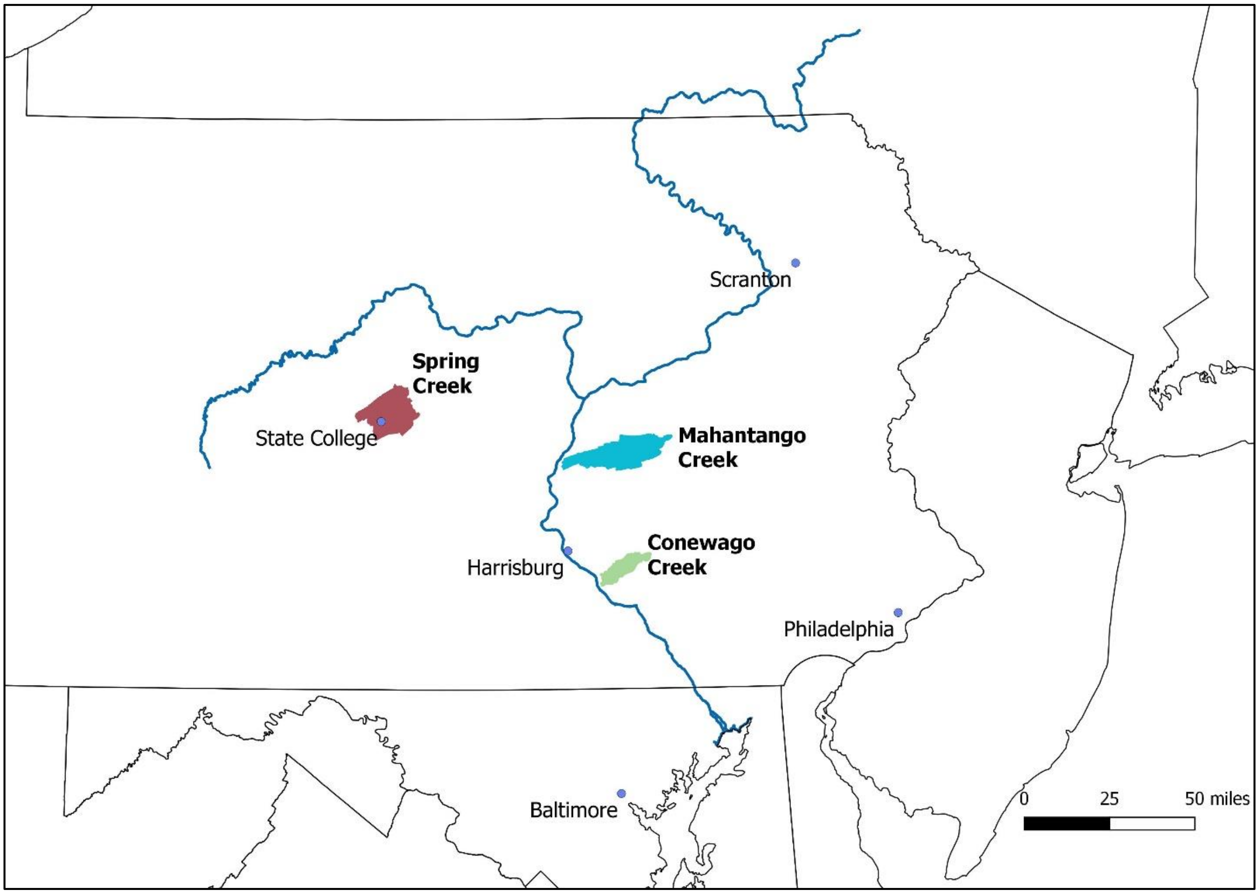

Figure 1. Location of target watersheds within the Susquehanna River basin

The Conewago Creek watershed is characterized by rolling farmland and forests, with some residential and commercial development. Water quality is designated as degraded in many stream segments due to excess nutrients and sediment. The Mahantango Creek watershed is lightly populated, with land cover consisting of forested ridges and valleys dominated by agricultural uses. Soil erosion is a major problem, as is unrestricted access of livestock to streams (Pennsylvania DEP. 2013). Spring Creek is geographically and geologically unique, characterized by a series of long, high ridges and broad limestone valleys. It is home to a number of rare, threatened, or endangered plants and animals and harbors one of the most productive wild trout fisheries in Pennsylvania. However, the Spring Creek watershed is experiencing rapid growth and development, which has resulted in degradation due to increased stormwater runoff, sedimentation, and reduced riparian protection (Biologists Reports PFBC 2007-2008).

For each stream segment in each target watershed, baseline values of the WQI were calculated using six water quality parameters: total nitrogen (N), total phosphorous (P), total sediments, dissolved oxygen (DO), biological oxygen demand (BOD), and fecal coliforms. N, P and sediment values for each stream segment were modeled using the SWAT model. DO, BOD and fecal coliforms were modeled using regressions that relate each of the three to N, P and Sediment. Weights for the six parameters were the same as those used by the EPA (EPA 2009a, EPA 2015). Average WQI values for small streams (stream order 2 and smaller) and larger streams (stream order 3 and larger) are presented in Table 1.[2]

		Stream Miles		Average WQI	
Watershed	Area (sq. mi.)	Small Streams	Large Streams	Small Streams	Large Streams
Spring Creek	144	29.5	20.2	59	60
Mahantango Creek	165	39.6	39.5	47	46
Conewago Creek	53	14.4	8.3	54	55

Table 1. Characteristics of the Study Watersheds

3.2. Survey Method and Design

Each survey respondent received a survey instrument that discussed only one of the three target watersheds.[3] For each target watershed, a sample of households was generated that lived within 50 miles of the boundary of the target watershed. Household addresses were purchased from a commercial source.

[2] Stream order (SO) is a way to classify streams by relative size. The smallest headwater streams are SO 1. When two SO 1 streams meet, they form a SO 2 stream. When two SO 2 streams meet, they form an SO 3 stream and so on.

[3] An example of the survey instrument is available as supplementary material.

A total of 1205 households were sampled across the three target watersheds. Households located closer to each watershed were oversampled, with 40% of the sample located within 15 miles of the watershed, 30% located between 15 and 30 miles from the watershed, and 30% located between 30 and 50 miles from the watershed. It should be noted that the sample areas for the three watersheds overlapped spatially, so that a household located close to one study watershed could receive a survey targeting a more distant watershed. Surveys were printed, mailed and optically scanned by the Pennsylvania State University Survey Research Center. Households received an initial survey packet with a $2 incentive, a follow-up postcard after one week, and a reminder packet with a replacement survey after three weeks (non-respondents only). Respondents were also given the opportunity to reply through a web version of the survey, though very few took advantage of that option.

Respondents were provided a map showing the boundaries of the target watershed and a description of the streams located within the target watershed, and were told the current average WQI value for small streams and large streams in the target watershed. A program was described to improve water quality in the target watershed by paying landowners to adopt best management practices. Respondents were told that the program could be designed to target small or large streams,[4] and could be more or less aggressive in its goals. Cost to the household was through higher taxes. The type of tax was not specified.

The attributes of each program were the percent of small streams in the target watershed that would be improved, the resulting WQI for small streams that were improved, the percent of large streams in the target watershed that would be improved, the resulting WQI for large streams that were improved, and the annual cost to the household. The attribute levels used in the experimental design are presented in Table 2.

Attributes	Levels

[4] Respondents were told that a small stream is one less than 15 feet wide, based on data on SO 2 and SO3 streams from Downing et al. (2012).

Percent of Small Streams Improved	**0%**, 50%, 100%
Water Quality Index in Improved Small Streams	**59**, 62, 65 (Spring Creek) **47**, 50, 53 (Mahantango) **54**, 57, 60 (Conewago)
Percent of Larger Streams Improved	**0%**, 50%, 100%
Water Quality Index in Improved Large Streams	**60**, 63, 66 (Spring Creek) **46**, 49, 52 (Mahantango) **55**, 58, 61 (Conewago)
Annual Cost to Household	**0**-54 dollars per year

Table 2. Attribute levels used in choice experiment questions. Baseline values are in bold.

It should be noted that the changes in the WQI valued here are smaller than those valued in most previous studies. It is common to value a change in quality from boatable to fishable (an increase of 25 WQI points) or from fishable to swimmable (an increase of 20 WQI points). The increases considered here were judged to be attainable using agricultural best management practices.

Each respondent answered eight choice questions. An example choice question is given in Figure 2. Each question presents a status quo option with baseline attribute levels and zero cost and two policy options with some or all of the attributes improving and positive costs. Respondents were asked to choose their preferred option. For each target watershed, two survey versions were created that differed in their experimental design.

	Program A	Program B	No Program
Percent of Small Streams Improved	100%	50%	0 %
Water Quality Index in Small Streams	65 points	62 points	59 points
Percent of Larger Streams Improved	100%	0%	0 %
Water Quality Index in Larger Streams	66 points	60 points	60 points
Cost to Your Household Each Year	$54/year	$18/year	$0/year
Your Choice	□	□	□

Figure 2. Example choice question

3.3. Econometric Modeling

Choices were modeled using a conditional logit model with a nonlinear utility function. The conditional logit model assumes that respondents assign a utility value to each option in a choice, according to some utility function, so that utility to respondent *j* from option *i* is given by

$$V_i = U(Miles_{SM}, Miles_{LG}, \Delta WQI_{SM}, \Delta WQI_{LG}, Cost, Dist) + \varepsilon_{ij}$$

where $Miles_{SM}$ and $Miles_{LG}$ are the number of miles of small and large streams improved under option *i* (scale), ΔWQI_{SM} and ΔWQI_{LG} are the improvement in the WQI in small and large streams that were improved (scope), *Cost* is the annual cost to the household of option i, *Dist* is the linear distance, in miles, between the centroid of the watershed being improved and respondent *j*'s zip code centroid, and ε_{ij} is an i.i.d. random term assumed to be distributed Type I Extreme Value. For each choice question, respondents are assumed to choose the option that gives the highest utility.

While it is common in choice experiments to use a linear form for the utility function, U, a linear form is not appropriate for this context. A water quality improvement that affects zero stream miles should have no value. Similarly, an improvement that results in zero change in the WQI should have no value, regardless of the number of stream miles. The good being valued is the combination of the WQI improvement and the number of stream miles improved. This suggests a multiplicative form. Further, it is reasonable to assume that all values will vary with distance, though the functional form of that distance decay is unknown. A utility function that satisfies these constraints is

$$U = (1 + \beta_d Dist + \beta_{d2} Dist^2) * (\beta_{SM} * Miles_{SM} * \Delta WQI_{SM} + \beta_{LG} * Miles_{LG} * \Delta WQI_{LG}) - \beta_C * Cost$$

The first part of the utility function accounts for distance decay in WTP, modeled as a quadratic form. Nonlinearity in distance decay can be examined by testing whether β_{d2}=0. The distance decay term is multiplied by a second term that incorporates miles improved and the scope of the improvement for both small and large streams. This form allows the value of improvements in large streams to differ from the value of improvements in small streams. This functional form imposes proportional relationships between WTP and both stream miles improved and the scope of the change in the WQI. In subsequent regressions, nonlinear forms are explored to investigate possible nonlinearities in scale and/or scope.

Estimation was done by maximizing the likelihood function using the LIMDEP Maximize routine.

4. Results

The overall response rate was 29%. It was noticed that the response rate decreased slightly with distance from the watershed, suggesting that the probability of response was negatively correlated with the value the respondent placed on the resource. To mitigate the potential for non-response bias, we adopted the conservative approach (Johnston et al., 2017, Morrison 2000) of assuming that all non-respondents have zero value for water quality improvements. Consequently, we assume that, had they responded, they would have chosen the baseline option in all questions. This approach allows inclusion of all respondents in the model estimation, but precludes the use of respondent-specific information in the model (because that information is not available for non-respondents). As a robustness check, additional models are estimated using only respondents.

4.1. Distance Decay

An initial conditional logit model was run to explore the shape of the distance decay function. For each watershed, sampled households were broken out into 10-mile distance bins, and model was estimated using dummy variables for each distance bin, with the coefficient for the 0-10 mile bin normalized to 1. This model included 9,640 choices from 1,205 sampled households. Figure 3 shows the values of the coefficient for each distance bin. These values represent the relative WTP for a household located within each distance bin, as a percentage of the WTP for households located in the 0-10 mile bin.

Values for every bin beyond 10 miles were significantly lower than the value for the 0-10 bin at the 1% significance level.

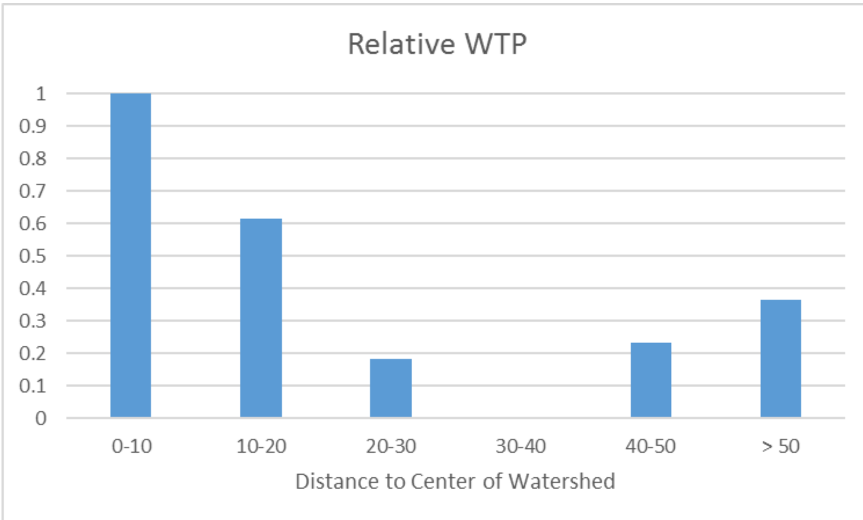

Figure 3. Relative WTP by distance to center of improved watershed

The pattern shown in Figure 3 suggests that values for water quality improvements decline with increasing distance from the targeted watershed out to around 40 miles, but then increase again past 40 miles. Why would values turn back upward past 40 miles? It is possible that respondents who live close to the targeted watershed focus only on that watershed when making choices, while respondents who live farther away value an improvement that affects a larger area. This is called part-whole bias, where the survey attempts to value an improvement to only part of the resource, while respondents value an improvement to the whole resource (Mitchell and Carson 1989).[5]

Respondents were specifically told that the program would only improve the targeted watershed. However, some respondents may have missed that statement or not believed it, or may believe that choosing programs that improve water quality in one watershed will increase the chance that similar

[5] Mitchell and Carson (1989) provide as an example of part-whole bias a respondent who is asked to value an improvement to a local river basin, but is unable to isolate only that basin in their mind, and "values a larger range of waters than intended."

programs will be adopted for other watershed. If part-whole bias does exist, and is more prevalent among respondents who live farther from the targeted watershed, it could account for the increase in WTP farther from the targeted watershed. To the extent that part-whole bias exists, it complicates our efforts to identify distance decay in values. As a conservative measure, we exclude from further analysis respondents who live more than 40 miles from the centroid of their target watershed. This left 6,592 choices from 854 potential respondents.

Using only respondents located within 40 miles of the centroid of their targeted watershed, the parameters of the utility function were estimated for both a quadratic distance decay function and a linear function with β_{d2} constrained to equal 0. Estimation results are presented in Table 3.

Coefficient	Linear Distance Decay	Nonlinear Distance Decay
β_{SM}	0.00691**	0.00804**
	(0.00105)	(0.00126)
β_{LG}	0.00970**	0.01144**
	(0.00142)	(0.00188)
β_{d}	-0.03031**	-0.04526**
	(0.00149)	(0.00659)
β_{d2}		0.00046*
		(0.00020)
β_{C}	-0.08817**	-0.08834**
	(0.00186)	(0.00186)
AIC	6530.0	6528.5

Table 3. Linear and nonlinear distance decay models (Standard errors in parentheses; * = p<0.05, ** = p<0.01)

The estimation results show that the distance decay function is nonlinear, with the squared distance term significantly different from zero at the 5% level. However, the practical implications of this nonlinearity may be small. Figure 4 shows estimated WTP for a 1 point increase in 1 mile of small and one mile of large streams, using both the linear and nonlinear distance decay functions. Both forms suggest that WTP drops to zero at a distance of 33 miles from the center of the watershed.[6] The difference in WTP between the linear and nonlinear forms at any given distance is small, suggesting that using the linear form (for convenience) would not result in large errors. Further, given our concern that results may be contaminated by part-whole bias, the evidence for nonlinearity in distance decay is weak at best.

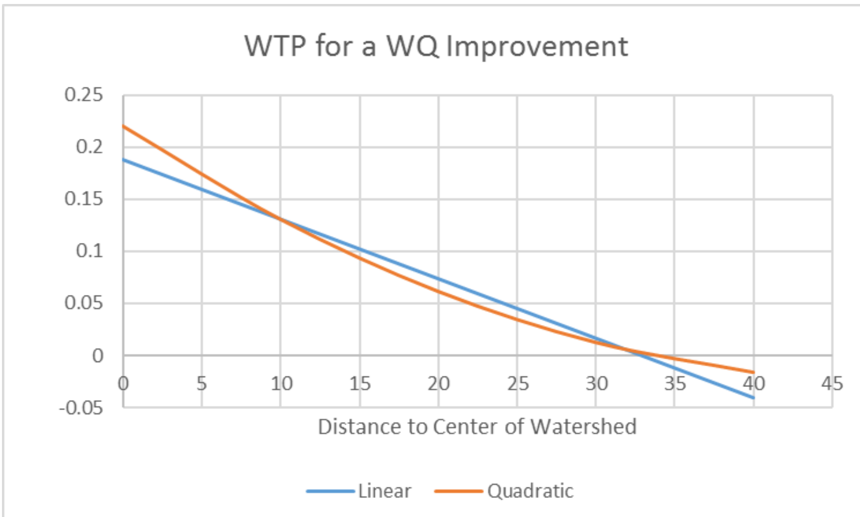

Figure 4. Linear and nonlinear distance decay functions

4.2. Large vs small streams

In both the linear and the nonlinear distance models, the coefficient on large stream improvements, β_{LG}, is around 40% larger than the coefficient on small stream improvements, β_{SM}. However, in neither model is this difference statistically significant[7], and we cannot reject a null hypothesis that $\beta_{LG} = \beta_{SM}$. But the precision with which the difference, $\beta_{LG} - \beta_{SM}$, is estimated is low. We also cannot reject a null hypothesis that $\beta_{LG} = 2*\beta_{SM}$. Our experimental design allows statistical identification of both parameters simultaneously, but is not powerful enough to precisely determine

[6] Regressions using only respondents showed slightly more curvature in the nonlinear form, but a very similar linear form with a zero-threshold distance of 34 miles.

[7] The same result held for a regression that used only respondents. Regressions run to test differences in parameters available from the authors.

whether they differ in size. We conclude that while our results show some evidence that large streams are valued more than small streams, that evidence is not conclusive. Absent strong statistical evidence that large streams are worth more than small streams, in subsequent models we restrict them to be equal, and create new variables $Miles_{ALL}$ and ΔWQI_{ALL} that measure the total miles of streams (large and small) improved and the average improvement in all improved streams.

4.3. Nonlinearities in Stream Miles and ΔWQI

The model used here imposes the assumption that respondents get utility from the product of the scale of the improvement (miles of streams improved) and the scope (size of the increase in WQI), i.e. $MilesAll_i$*$\Delta WQIAll_i$. Before exploring whether either the scale or the scope might enter into the utility function in nonlinear way, it is useful to confirm that the functional form is appropriate. In particular, we need to establish that respondents are reacting to both scale and scope when choosing an option. This was tested by estimating two new models (using the linear distance decay function) as follows

$$U = (1 + \beta_d * Dist) * \left(\beta_{Mi} * Miles_{ALL} + \beta_{Mi*WQI} * Miles_{ALL} * \Delta WQI_{ALL}\right) - \beta_C * Cost$$

and

$$U = (1 + \beta_d * Dist) * \left(\beta_{WQI} * \Delta WQI_{ALL} + \beta_{Mi*WQI} * Miles_{ALL} * \Delta WQI_{ALL}\right) - \beta_C * Cost$$

In the first model, if β_{Mi} is nonzero and β_{Mi*WQI} equals zero, then respondents are reacting only to stream miles, and not to ΔWQI. In the second model, if β_{WQI} is nonzero and β_{Mi*WQI} equals zero, then respondents are reacting only to ΔWQI, and not to the number of stream miles improved. In the first estimation, β_{Mi*WQI} was significantly positive while β_{Mi} was not significantly different from zero, indicating that respondents do react to both scale and scope, and not scale alone. In the second, β_{Mi*WQI}

was significantly positive while β_{WQI} was not significantly different from zero, indicating that respondents, again, do react to both scale and scope, and not scope alone.

To test whether either scale or scope might enter into the interaction in a nonlinear way, we estimate two new models (again using the linear distance decay function)

$$U = (1 + \beta_d * Dist) * \left(\beta_{Mi1} * Miles_{ALL} * \Delta WQI_{ALL} + \beta_{Mi2} * {Miles_{ALL}}^2 * \Delta WQI_{ALL}\right) - \beta_C * Cost$$

and

$$U = (1 + \beta_d * Dist) * \left(\beta_{WQ1} * Miles_{ALL} * \Delta WQI_{ALL} + \beta_{WQ2} * Miles_{ALL} * {\Delta WQI_{ALL}}^2\right) - \beta_C * Cost$$

In neither model was the nonlinear term statistically different from zero at the 5% level. We find no evidence that WTP is nonlinear in either the number of stream miles improved or the scope of the water quality improvement. Regressions using only respondents gave the same result.

Because only three target watersheds were included in this study, we do not have enough variation in baseline conditions to reliably determine whether WTP varies with baseline water quality.

4. *WTP for Water Quality Improvements*

We estimate a value function that assumes WTP is proportional to both scale and scope. This function is based on the following parsimonious utility model

$$U = (1 + \beta_d * Dist) * (\beta_{All} * Miles_{ALL} * \Delta WQI_{ALL}) - \beta_C * Cost$$

Parameter estimates from an estimation using all household located within 40 miles of the centroid of the watershed are presented in Table 4.

Coefficient	Parameter Estimate
β_{ALL}	0.00791** (0.00062)
β_d	-0.03048** (0.00155)
β_C	-0.08795** (0.00188)
AIC	6530.0

Table 4. Choice model (Standard errors in parentheses; * = p<0.05, ** = p<0.01)

This model generates a value function for an improvement to a specific stream of the form

$$WTP = (1 - 0.03048 * Dist) * 0.0899 * Miles * \ \Delta WQI$$

where *Dist* is the linear distance (in miles) from the household to the center point of the improved stream, *Miles* is the number of stream miles improved, and ΔWQI is the scope of the improvement for that stream. Programs that improve multiple streams can be valued by calculating the WTP for each improved stream, and summing.

5. Discussion

The findings of this study have important implications for cost-benefit analyses of specific water quality interventions as well as for regulatory impact analyses of nation-wide regulations.

Results from this study indicate that, for water quality changes of a scope that might result from adoption of agricultural best management practices or from national regulations under the Clean Water Act, there is little evidence that distance decay is nonlinear in an important way, though part-whole bias complicates this conclusion. There is weak evidence that large rivers are worth more than small rivers. There is no evidence that WTP is nonlinear with the degree to which the quality of the water has been improved, nor how many stream miles have been improved. These last results, that values for surface water quality improvements are proportional to the scope and scale of the improvements, imply that it is defensible to value individual interventions independently and then sum them altogether.

These results validate some, but not all, of EPA's practices when conducting national-level regulatory impact assessments. Specifically, our results suggest that the practice of calculating a WTP per unit change in WQI and then multiplying that unit value by projected changes in WQI for stream segments, as was done by the EPA in their 2015 Electric Power RIA, is defensible, at least for small increases in the WQI. Further, our results do not clearly reject EPA's practice of valuing streams of different size (stream order) equally, though more research is needed to more-precisely determine whether larger streams are worth more to households than smaller streams.

However, our results call into question EPA's treatment of distance. Our ability to determine whether distance decay is linear or nonlinear was complicated by potential part-whole bias. But regardless of the functional form, the results clearly show that closer streams are more highly valued than more distant streams. Further, our analysis, as well as previous literature, suggests that WTP for water quality improvements decreases to zero at a distance less than the 100 mile cutoff used in EPA's 2015 Steam Electric RIA. EPA's approach in their 2009 C&D RIA, which assumed that all rivers and streams in the household's home state are equally valued, is clearly not supported by the evidence.

It is important to note that the water quality improvements valued here are relatively modest. They were chosen to reflect improvements that could be achieved through adoption of agricultural best management practices. It may well be that important nonlinearities exist when considering more dramatic water quality improvements.

We stress three caveats with our results that call for future research. First, our survey appears to have suffered from part-whole bias, particularly for respondents located farther from valued watersheds. We explicitly told respondents that the water quality improvement would only affect that watershed that was described to them, but it appears that some respondents valued a program that benefited a larger area. Second, our results regarding small vs large streams lack precision. We cannot reject the null hypothesis that small and large streams are equally valued, but we can also not reject a null hypothesis that large streams are worth twice as much as small streams. Our experimental design was simply not powerful enough to identify the difference in the two values with precision. Third, our survey included only three watersheds, which is not enough to reliably estimate any possible relationship between baseline water quality and WTP for a water quality improvement over the baseline. Although our three watersheds do vary in their baseline water quality, they also vary in other attributes as well. To reliably determine the relationship between WTP for an improvement in water quality and baseline water quality would require a survey effort that includes a much larger and more diverse set of target watersheds.

References

Alpizar Rodriguez, Francisco, Fredrik Carlsson, and Peter Martinsson. "Using choice experiments for non-market valuation." No. 52. University of Gothenburg, Department of Economics (2001).

Bateman, Ian, Alistair Munro, Bruce Rhodes, Chris Starmer, and Robert Sugden. "Does part–whole bias exist? An experimental investigation." The Economic Journal 107.441 (1997): 322-332.

Bateman, Ian J., Ian Langford, Naohito Nishikawa, and Iain Lake. "The Axford debate revisited: a case study illustrating different approaches to the aggregation of benefits data." (2000): 291-302.

Biologist Reports PFBC. "Dam Removal, Habitat Improvement, and Trout Surveys (2007-2008): Spring Creek Centre County"

Boyle, Kevin J., William H. Desvousges, F. Reed Johnson, Richard W. Dunford, and Sara P. Hudson. "An investigation of part-whole biases in contingent-valuation studies." Journal of Environmental Economics and Management 27.1 (1994): 64-83.

Carson, Richard T., and Robert Cameron Mitchell. "The value of clean water: the public's willingness to pay for boatable, fishable, and swimmable quality water." Water resources research 29.7 (1993): 2445-2454.

Downing, J.A., J.J. Cole, C.M. Duarte, J.J. Middelburg, J.M. Melack, Y.T. Prairie, P. Kortelainen, R.G. Striegl, W.H. McDowell, and L.J. Tranvik. "Global abundance and size distribution of streams and rivers." Inland Waters 2(2012):229-236.

Economics for the Environment Consultancy (Eftec). "Valuing Environmental Impacts: Practical Guidelines for the Use of Value Transfer in Policy and Project Appraisal: Technical Report" Department for Environment, Food and Rural Affairs, U.K. (2010).

Economics for the Environment Consultancy (Eftec). "Valuing Environmental Impacts: Practical Guidelines for the Use of Value Transfer in Policy and Project Appraisal: Value Transfer Guidelines" Department for Environment, Food and Rural Affairs, U.K. (2009).

E.P.A. (U.S. Environmental Protection Agency) "Environmental and Economic Benefit Analysis of Final Revisions to the National Pollutant Discharge Elimination System Regulation and the Effluent Guidelines for Concentrated Animal Feeding Operations (CAFO)". EPA-821-R-03-003. December (2002)

E.P.A. (U.S. Environmental Protection Agency) "Environmental Impact and Benefits Assessment for Final Effluent Guidelines and Standards for the Construction and Development Category." EPA-821-R-09-012. November (2009a)

E.P.A. (U.S. Environmental Protection Agency) "Economic Analysis of Final Effluent Limitation Guidelines and Standards for the Construction and Development Industry." EPA-821-R-09-011. November (2009b)

E.P.A. (U.S. Environmental Protection Agency) "Benefit and Cost for the Effluent Limitations Guidelines and Standards for the Steam and Electric Power Point Source Category." EPA-821-R-15-005. November (2015)

Farber, Stephen, and Brian Griner. "Valuing watershed quality improvements using conjoint analysis." Ecological Economics 34.1 (2000): 63-76.

Foster, Vivien, and Susana Mourato. "Elicitation format and sensitivity to scope." Environmental and resource economics 24.2 (2003): 141-160.

Ge, Jiaqi, Catherine L. Kling, and Joseph A. Herriges. "How much is clean water worth? Valuing water quality improvement using a meta analysis." (2013).

Georgiou, Stavros, Ian Bateman, Matthew Cole and David Hadley. "Contingent ranking and valuation of river water quality improvements: Testing for scope sensitivity, ordering and distance decay effects." Centre for Social and Economic Research on the Global Environment Working Paper 2000-18, (2000).

Goldberg, Isabell, and Jutta Roosen. "Scope insensitivity in health risk reduction studies: A comparison of choice experiments and the contingent valuation method for valuing safer food." Journal of Risk and Uncertainty 34.2 (2007): 123-144.

Griffiths, Charles, Heather Klemick, Matt Massey, Chris Moore, Steve Newbold, David Simpson, Patrick Walsh, and William Wheeler. "US Environmental Protection Agency valuation of surface water quality improvements." Review of Environmental Economics and Policy (2012): rer025.

Hanley, Nick, Robert E. Wright, and Begona Alvarez-Farizo. "Estimating the economic value of improvements in river ecology using choice experiments: an application to the water framework directive." Journal of environmental management 78.2 (2006): 183-193.

Hanley, Nick, Felix Schläpfer, and James Spurgeon. "Aggregating the benefits of environmental improvements: distance-decay functions for use and non-use values." Journal of environmental management 68.3 (2003): 297-304.

Hanley, Nick, Susana Mourato, and Robert E. Wright. "Choice modelling approaches: a superior alternative for environmental valuation?" Journal of economic surveys 15.3 (2001): 435-462.

Heisler, John, P. Glibert, J. Burkholder, D. Anderson, W. Cochlan, W. Dennison, C. Gobler, Q Dortch, C. Heil, E. Humphries, A. Lewitus, R. Magnien, H. Marshall, K. Sellner, D. Stockwell, D. Stoecker, and M. Suddleson. "Eutrophication and harmful algal blooms: a scientific consensus." Harmful Algae 8.1 (2008): 3-13.

Hite, Diane, Darren Hudson, and Walaiporn Intarapapong. "Willingness to pay for water quality improvements: The case of precision application technology." Journal of Agricultural and Resource Economics (2002): 433-449.

Hoehn, John P., and Alan Randall. “Too Many Proposals Pass the Benefit Cost Test.” The American Economic Review 79.3 (1989):544-551.

Johnston, R.J., K.J. Boyle, W. Adamowicz, J. Bennett, R. Brouwer, T.A. Cameron, W.M. Hanemann, N. Hanley, M. Ryan, R. SCarpa, R. Tourangeau, and C.A. Vossler. “Contemporary guidance for stated preference studies.” Journal of Association of Environmental and Resource Economists 4(2017):319-405.

Kim, J.H., S.N. Kim, and S. Doh. “The distance decay of willingness to pay and the spatial distribution of benefits and costs for the ecological restoration of an urban branch cream in Ulsan, South Korea.” Annals of Regional Science 54(2015):835-853.

Kling, C.L., and D.J. Phaneuf. How are scope and adding up relevant for benefits transfer? Environmental and Resource Economics 69(108):483-502.

Loomis, J.B. “How large is the extent of the market for public goods: evidence from a nationwide contingent valuation survey.” Applied Economics 28(1996):779-782.

Loomis, John, Paula Kent, Liz Strange, Kurt Fausch, and Alan Covich. "Measuring the total economic value of restoring ecosystem services in an impaired river basin: results from a contingent valuation survey." Ecological economics 33.1 (2000): 103-117.

Lumb A., T.C. Sharma, and Jean-Francois Bibeault. "A review of genesis and evolution of water quality index (WQI) and some future directions." Water Qual Expo Health 3(1) (2011):11–24

Mitchell, R. C. and R. T. Carson, "Using Surveys to Value the Benefits for Public Goods." Resources for the Future, Washington, D.C. (1989)

Morrison, M. “Aggregation bias in stated preference studies.” Australian Economic Papers 39(2000):215-230.

Parsons, G.R., “Travel Cost Models.” In P.A. Champ, K. Boyle and T.C. Brown (eds) A Primer on Nonmarket Valuation, 2nd Edition. Springer Nature, Dordrecht, The Netherlands. 2017.

Pennsylvania DEP (Department of Environmental Protection) "Mahantango Creek Subwatershed TMDL: Northumberland and Schuylkill Counties" March 27 (2013)

Smith, V.K., and W.H. Desvouges. Measuring water quality benefits. Boston: Kluwer Nijhoff. 1986.

Tyagi, Shweta, Bhavtosh Sharma, and Prashant Singh. "Water quality assessment in terms of water quality index." American Journal of Water Resources 1.3 (2013): 34-38.

Vajjhala, S.P., A.M. John, and D.A. Evans. “Determining the extent of the market and extent of resource for stated preference survey design using mapping methods.” EPA NCEE Working Paper #08-09. Washington, DC. 2008.

Van Houtven, George, John Powers, and Subhrendu K. Pattanayak. "Valuing water quality improvements in the United States using meta-analysis: Is the glass half-full or half-empty for national policy analysis?." Resource and Energy Economics 29.3 (2007): 206-228.

Vaughan, W.J. “The water quality ladder.” In R.C. Mitchell and R.T. Carson (appendix II), An experiment in determining willingness to pay for national water quality improvement, draft report. (Available at: https://www.epa.gov/sites/production/files/2017-12/documents/ee-0011_1-5.pdf).

Walsh, P.J., and W.J. Wheeler. Water quality indices and benefit-cost analysis. Journal of Benefit Cost Analysis 4(2013):81-105.