PNAS

Supplementary Information for
Valuing Improvements in Ecological Integrity in Local and Regional Watersheds: The Biological Condition Gradient Ladder

Christian A. Vossler, Christine L. Dolph, Jacques C. Finlay, David A. Keiser, Catherine L. Kling, Daniel Phaneuf

Daniel Phaneuf
Email: dphaneuf@wisc.edu

This PDF file includes:

Supplementary text
Figures S1 to S3
Tables S1 to S7
SI References

Supplementary Information Text

S1. BCG Protocols

Conversion of biological integrity to BCG scores and spatial transferability of biological condition estimates. At the time of data acquisition, state agency personnel in four states (IL, IN, MN, OH) had developed BCG scoring criteria for streams and rivers according to the BCG framework outlined in USEPA (2016; Table S1). Although the remaining states did not have BCG criteria explicitly developed, each had index of biological integrity (IBI) scores on which stream condition was evaluated. In order to develop a high-resolution estimate of biological condition across the entire river basin, we sought to convert the biological index scores used by states without a BCG to 'BCG proxies'. Previous analysis has shown IBI scores to be strongly related to BCG levels in multiple states (USEPA 2016; Bouchard, 2016). Most states summarize IBI scores into narrative categories that can be used to describe stream condition – i.e., data are used to classify sites into categories of 'excellent', 'good', 'fair', 'poor', based on their IBI score. These narrative biological categories were used to generate an approximate estimate of biological condition that is spatially transferable across our entire study area. Narrative biological categories used in each state were converted to a BCG proxy using both the meaning conveyed by each category together with what we know about BCG scores, based on the states where both BCG and multi-metric index criteria have been developed. For example, BCG Level 1 sites for macroinvertebrate communities are exceedingly rare in developed regions (Gerritsen and Stamp, 2016). In Midwestern states where BCG scores have been developed (i.e., MN, IL, IN, OH), there are no sites in the UMOHTN that have been assessed as BCG Level 1. Based on the assumption that Level 1 sites were very unlikely in our study region, narrative biological categories used by states were assigned a BCG proxy level of between 2 and 6, with one exception. The state of North Carolina had a narrative category defined as 'Natural'. Given that North Carolina is characterized by considerable extents of undeveloped and forested areas, we assigned this narrative category to be Level 1 on the BCG scale. However, none of the Level 1 sites from North Carolina occurred within the UMOHTN.

In several states, analyses by state agencies have identified the impairment threshold for streams and rivers to fall typically between BCG level 3 and 4 (USEPA, 2016). We used this criterion to help translate narrative categories used by states into BCG proxies. Where states had fewer than 5 narrative categories, we split narrative categories such that unimpaired sites were split across levels 2 and 3, ambiguously impaired sites were split across levels 3 and 4, and severely impaired sites were split across levels 5 and 6. In states with more than five narrative categories, we assigned sites that were in very good condition to Level 2, good condition to Level 3, sites in fair to moderately stressed condition to Level 4, sites in poor condition to Level 5, and sites that were in very poor condition to Level 6. Table S1 presents biological index score ranges and narrative categories that were associated with each 'BCG proxy' level for each state.

Although we sought to create estimates of biological condition that were transferable across state boundaries from the biological indices available, the indices were not all created in a fully standardized manner across different states. Our analysis is thus potentially limited by discrepancies in how biological condition was measured and portrayed across state boundaries (USEPA, 2016). However, our estimates of BCG condition based on closely related biological index scores (Bouchard 2016) represented the best available method for generating a single measurement of biological condition across the entire study region and are likely to accurately reflect the approximate biological condition of streams and rivers at the coarser scales we used for this study (HUC8 and HUC4 watersheds).

Overlaying use categories on BCG levels. Because we did not have direct information (i.e., empirical studies) linking different BCG levels to stream and river uses (such as swimming, fishing, boating), we used our knowledge of specific water bodies in the Midwest together with their BCG levels (Figure S1) to assign use categories to each level more broadly across the study basin. For example, the St. Croix River forming the border of Minnesota and Wisconsin is designated by the National Park Service as a national scenic riverway, and (at the time of this study), was categorized largely by BCG Level 2 in the upper reaches using BCG levels based on fish community data. This river has many beaches that are popular for swimming, particularly in the upper portion of the watershed, and is widely used for both boating and fishing. Based on our knowledge of this river, we assigned Level 2 BCG sites the uses of swimming, wading, fishing, and boating (all common uses for this river). We then assumed that all other Level 2 streams and rivers offer similar opportunities for recreational use. In other words, we assumed that the conditions that made the St. Croix River suitable for all recreational uses would also apply to other level 2 ecosystems, and we assigned all of these uses to BCG level 2 streams and rivers. By comparison, the Mississippi River near St. Cloud, Minnesota, is categorized largely as BCG level 3. While the Mississippi River near this location is commonly used for boating and fishing and even wading, it is less common for people to swim in the river. Thus, we assigned this and all other level 3 rivers the uses of wading, fishing, and boating. Many Level 4 rivers, such as the North Fork of the Crow River, have considerably higher turbidity and algal growth (MPCA, 2021), and are less likely to be used for swimming or wading. However, these rivers are likely quite productive for some game fish since river production can increase at moderate levels of eutrophication despite other disturbances (Finlay 2011). Thus, we included fishing and boating as uses for Level 4 systems, but not swimming or wading. Rivers like the South Fork of the Crow River, characterized largely by BCG Level 5, are even more turbid and are more likely to be characterized by high growth of bacteria (MPCA, 2016), and are therefore unlikely to be used for swimming or wading. Some game fish (such as catfish, for example) may still be common under Level 5 conditions; however, the fish community as a whole is likely to be strongly impacted. Thus, for Level 5 sites we included boating but not fishing as the primary use. Lastly, Level 6 rivers are the most highly degraded, are likely to be highly turbid with high levels of bacterial and algal growth, and harbor very few fish or insect species. Thus, we assumed that Level 6 streams and rivers would generally not be appropriate for any recreational use.

Once we established how recreational use intersected with BCG levels for a subset of local rivers that we knew well, we generalized these relationships to the rest of the study region. That is, we assumed that the uses that applied to the rivers we knew would also apply to streams and rivers of the same BCG level across the study region. This process for overlaying uses on to BCG levels was not perfect or exact, and relied on authors' knowledge of conditions in part of the study watershed (for Minnesota in particular) that may not apply uniformly across the study area. Additional study is needed to determine how recreational or other uses correspond to biological condition in different regions and could help inform further study aimed at assessing total value of water quality based on use and nonuse values.

S2. Survey and Experimental Design

Additional survey details. The start of the online survey encouraged respondents to use a laptop, desktop, or large tablet computer, if possible, to better view the displayed graphics. This was followed by an authentication screen that asked for the respondent's residential zip code. Providing a valid zip code located in our study area was necessary for continuing the survey. This was followed by introductory screens that highlighted the intent of the study, provided researcher affiliations, and indicated the study was funded by the Environmental Protection Agency. The survey was divided into four parts. Part 1 asked for opinions on water resources and quality near the respondent's home. Part 2 described the impacts of pollution in lakes, rivers, and streams, and used images along with written descriptions to separately characterize the appearance, supporting uses, and biodiversity of waterbodies as delineated according to six water quality levels. While these six levels correspond with the Biological Condition Gradient (BCG), this terminology was not introduced. Instead, the six levels were labelled as "Level 1 – Natural State", "Level 2 – Close to Natural State, "Level 3 – Some Changes Noticeable", "Level 4 – Many Changes Noticeable", "Level 5 – Major Degradation", and "Level 6 – Extreme Degradation".

Graphics used to distinguish the six BCG levels, as illustrated in Figure 2, are each composed of three images separately characterizing appearance, supporting human uses, and biodiversity. To make sure that the images were understood, and as attention checks, we presented a set of three images and asked the respondent to select the image that corresponded with the best water quality. This was done separately using images characterizing appearance, supporting uses, and biodiversity. If the respondent answered incorrectly, they were then shown the correct answer along with a refresher on how to interpret the images.

Following the discussion of the six BCG levels, respondents were shown a map that displayed the spatial distribution of water quality (based on the six water quality levels) for a specific watershed that did not include the respondent's home. The respondent was then asked to locate a specific city on the map and identify the water quality level for that city according to the map. Overall, there were four questions used to gauge understanding. Respondents were removed from the estimation sample if they incorrectly answered more than two of them, and further completed the survey in less than 10 minutes. Part 2 of the survey concluded by displaying a map of the respondent's home watershed. The respondent was asked to identify water quality near their home.

Part 3 of the survey included the valuation scenarios. These scenarios were preceded by an overview of how a new policy would improve water quality, how a new policy would be paid for, and possible reasons why people might vote yes or no on a proposal. The survey then informed participants that they would vote on several proposals and instructed them to consider each proposal separately. That the survey results will be shared with public authorities was emphasized prior to voting. We randomly varied information scripts related to the payment scheme (three options) and how the authorities may interpret the results from the study (two options). While the effect of these scripts is a topic for future work, the scripts overall had only small effects (2 to 4 percentage point differences) on the three belief measures mentioned in the article that are related to the incentive compatibility of the survey.

Part 3 of the survey concluded with a set of follow-up questions designed to understand whether people voted in ways congruent with incentive compatible elicitation (e.g., whether they voted as if the survey would be used to inform policy); what attribute changes influenced votes; response uncertainty; beliefs about distributional aspects of the tax burden; and general thoughts about the water quality proposals. Part 4 was short, and mainly asked about water-based recreation activities, and how the COVID-19 pandemic may have impacted income and local water-based recreation activities. Demographic information was obtained from the survey research firm, rather than through the survey.

Additional experimental design details. Proposals were grouped into "spatial unit" blocks and "BCG change" blocks. The intent of this blocking design was to lessen the cognitive burden on respondents by reducing the number of attributes that varied across voting scenarios. A "spatial unit" block held fixed the spatial unit, defined by spatial scale and whether the scenario was local or non-local (there are five possibilities), while varying the BCG change scenario. As there are five spatial units, there are five blocks in the design. Within a block there are four proposals – one for each BCG change. A "BCG change" block held fixed the change scenario (four possibilities; see Table 1) while varying the spatial unit. There are four BCG change blocks, and within a block there are five voting proposals.

Each respondent was randomly assigned either two (of the five) spatial unit blocks or two (of the four) BCG change scenario blocks. The order of proposals within a block was randomized. This resulted in eight proposals for those selected in the spatial unit blocks and ten proposals for those selected in the BCG change scenario blocks. However, in some cases, selected proposals were not presented because they were either repetitive (e.g., if an entire policy area is at Level 3, then the 'minimum Level 2' and 'change Level 3 to Level 2' change scenarios are identical) or redundant (e.g., if an entire policy area is at Level 3, then the 'minimum Level 3' scenario is meaningless). Overall, the number of proposals voted on varied from six to ten.

Voting proposals that were "local" included information on current and proposed conditions for the respondent's home sub-watershed (8-digit HUC), as determined by the respondent zip code. While "local" spatial units were endogenous to the respondent, based on where they lived, the "non-local" spatial units were randomly assigned. Specifically, we randomly assigned each zip code to a watershed different from their home watershed. This watershed, and a three-watershed grouping that contained the randomly assigned non-local watershed, were used in voting proposals.

Each voting scenario began with an introductory statement:

> *"The water quality changes described below would occur only in the highlight policy region on the maps. Improvements would occur gradually, reaching the new conditions by about 2026, and then remain at the new levels. The tax increase would last 5 years, and be in place from 2022 to 2026."*

This was followed by color-coded maps illustrating current and proposed water quality conditions, as depicted in Figures 4. This was followed by clickable thumbnails of the graphics describing the six water quality levels that were introduced in Part 2, a tabular summary of the proposed policy, and an advisory referendum question that solicited a yes or no vote on the proposal. An example of the tabular summary is included as Figure S3.

S3. Econometric analysis of survey data

Econometric framework. The theoretical foundation for data from a discrete choice experiment such as ours is random utility maximization (RUM) theory, which assumes that the utility a person derives depends on observed characteristics of choices as well as sources of well-being unobserved by the analyst, represented by a random component (McFadden 1974). Let the utility of respondent *i* derived from option *j* in voting scenario *t* be given by

$$V_{ijt} = \alpha c_{ijt} + \mathbf{x}_{ijt}\boldsymbol{\beta} + u_{ijt}, \tag{1}$$

where utility V_{ijt} is assumed to be additively separable in the cost of the alternative, c_{ijt}, and other attributes, $\mathbf{x}_{ijt}$. The symbols α and $\boldsymbol{\beta}$ denote estimable parameters and u_{ijt} is a random error term assumed to be independent and identically distributed Type I extreme value.

A common estimator for (1) is the mixed logit, which extends the canonical conditional logit model by allowing one or more of the utility parameters to be random and follow parametric distributions (Revelt and

Train 1998). This accommodates multiple sources of unobserved variability, including heterogeneity in tastes across respondents. The willingness to pay (WTP) estimates reported in Table 2 and Table 3 are derived from two mixed logit models, which we refer to as Model 1 and Model 2, respectively. Both models are estimated using the “mixlogit” command in Stata version 17.0. Estimation was based on 500 Halton draws.

Table S4 presents Model 1 estimates. The model includes as covariates the first eight variables defined in Table S3. The main covariates are the BCG scores, specific to the spatial unit involved in the voting scenario. As such, a parameter measures the marginal utility associated with an increase in the BCG score, and marginal utilities are allowed to freely vary across different spatial units. For local policies, the model further captures changes in utility associated with changes in the BCG score specific to the sub-watershed (8-digit HUC) where the respondent lives. All utility parameters, except for the parameter on the cost variable, were assumed to follow normal distributions. Assuming a fixed cost coefficient is common when the goal of the model is to estimate WTP (see Greene 2018). For random parameters, the means and standard deviations of the normal distributions are presented.

Table S5 presents mixed logit results for Model 2. The specification is similar to the prior model, but the data are only from voting scenarios that focused on either a local or non-local (4-digit HUC) watershed. To quantify preferences for in-state versus out-of-state water quality improvements, we interacted the last two variables defined in Table S3 with the alternative-specific constant.

Within this estimation framework, the negative of the parameter on the cost variable, $-\alpha$, is a measure of the marginal utility of income. Dividing the parameter on a non-cost attribute k by the marginal utility of income, $-\beta_k/\alpha$, yields the marginal WTP for an increase in attribute k. To derive the (total) WTP measures reported in Table 2 and Table 3, we calculate the difference in utility between an improved scenario and current conditions and divide it by the marginal utility of income. This calculation is done separately for each respondent, and then averaged over the sample:

$$WTP^{improved} = \sum_{i=1}^{N}\{-(\mathbf{x}_i{}^{improved} - \mathbf{x}_i{}^{curent})\boldsymbol{\beta}/\alpha\} \quad (2)$$

To calculate the standard errors for WTP estimates, we use the delta method. This is carried out using the “margins” post-estimation command in Stata.

We note that basing estimation on the simpler conditional logit model, or a more complex mixed logit model that allows the utility parameters to be correlated, produced similar WTP estimates, and qualitatively similar results. Further, using post-stratification raking weights to control for small differences between the characteristics of our sample and those of the broader population has a negligible effect on estimates.

Accommodating distance and space. Table S3 defines the variables used in our mixed logit regressions. With these we accommodate distance and space in two ways. First, our BCG scores are interacted with indicators for the size of the policy area, so that we recover utility differences stemming from improvements in the BCG that are specific to the spatial scale of the policy (sub-watershed, watershed, group of three watersheds, and the full study area). These estimates allow us to understand the *spatial scale* effect, which we define as the additional WTP for expanding a policy area, while holding fixed the location of the household. Second, our policy scenarios are differentiated by local and non-local changes, and the BCG effect is separately estimated for the two. This allows us to differentiate the WTP for changes at the watershed and group of three watersheds based on whether the policy area includes the respondent’s residence. With the non-local estimates, we can understand the *extent of the market* effect, which we define as the geographical range of the population that receives economic benefits from an improvement, holding fixed the location and size of the policy area.

Our use of discrete variables to describe space and distance is different from many stated preference studies. In these applications, the resource to be valued is fixed at a point in space and so the distance from the resource to each respondent’s home can be computed and compared. In these experimental designs, continuous distance is used as a variable to estimate the “distance decay” effect, in which a typical finding is that people are willing to pay less for an environmental improvement farther from their home. This type of finding is useful because the distance variable measures can be measured and interpreted in a consistent way for all respondents.

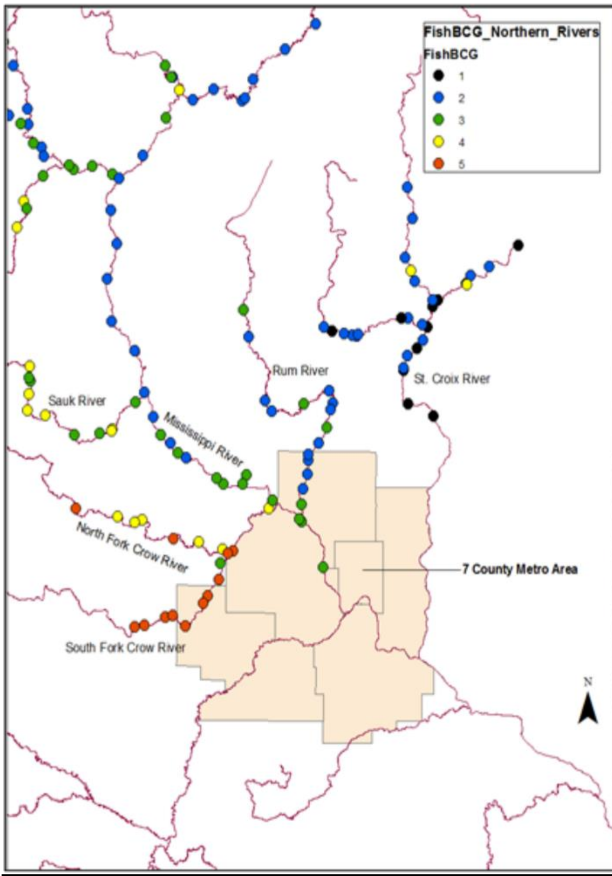
Our experimental design, however, uses varying-in-size spatial units spread throughout the study area, with a similarly dispersed sample. Distance measures become increasing challenging as the spatial scale increases, and it is no longer obvious how to define the endpoints of a distance measure. For respondents living in different parts of the study area, distances are relative to different baseline resources. For changes that occur throughout a local watershed, a respondent may care about distance to the center of the policy area, the distance to the changed area that is furthest away, the distance to the changed area where she recreates, or perhaps considers multiple locations within the watershed in which case an index that accounts for distance to multiple points may be warranted.

We have nonetheless examined specifications that use linear distance from the center of the respondent's zip code to the center of the policy area. The estimated utility changes based on this distance measure are virtually zero, and statistically insignificant. This evidence, in part, motivated us to explore the role of in-state percentage as an alternative geographic indicator. How to capture distance effects associated with changes across a large landscape is an important topic for future research.

Summary statistics and the effects of socio-economic characteristics. Table S6 summarizes socio-economic characteristics of the sample, along with variables obtained from the survey that are referred to in the article. While we utilized a probability sample, there are nevertheless some relatively minor differences between the sample and population of our study area. The sample is slightly better educated, older, and with higher incomes.

To explore whether WTP estimates vary by socio-economic characteristics, which is useful for meta-analyses and benefit-transfer, we estimated a mixed logit model that extends the specification of Model 1. Added to Model 1 are a set of interaction variables created by multiplying the socio-economic characteristics summarized in Table S6 with the alternative specific constant (ASC). The ASC captures the difference in utility between current conditions and *any* improvement scenario (regardless of the extent of the improvement). Dividing the coefficient on an interaction variable by the marginal utility of income produces a marginal WTP estimate.

Table S7 presents the marginal WTP estimates from the extended model (estimates associated with policy attributes are omitted for brevity). As indicated, WTP decreases with age, and increases with education and income. The marginal WTP associated with income suggests that, for every $1000 increase in annual income, WTP for any improvement policy increases by 72 cents. Identifying as female, retired, living in a metro area, household size and race/ethnicity have insignificant effects on WTP.



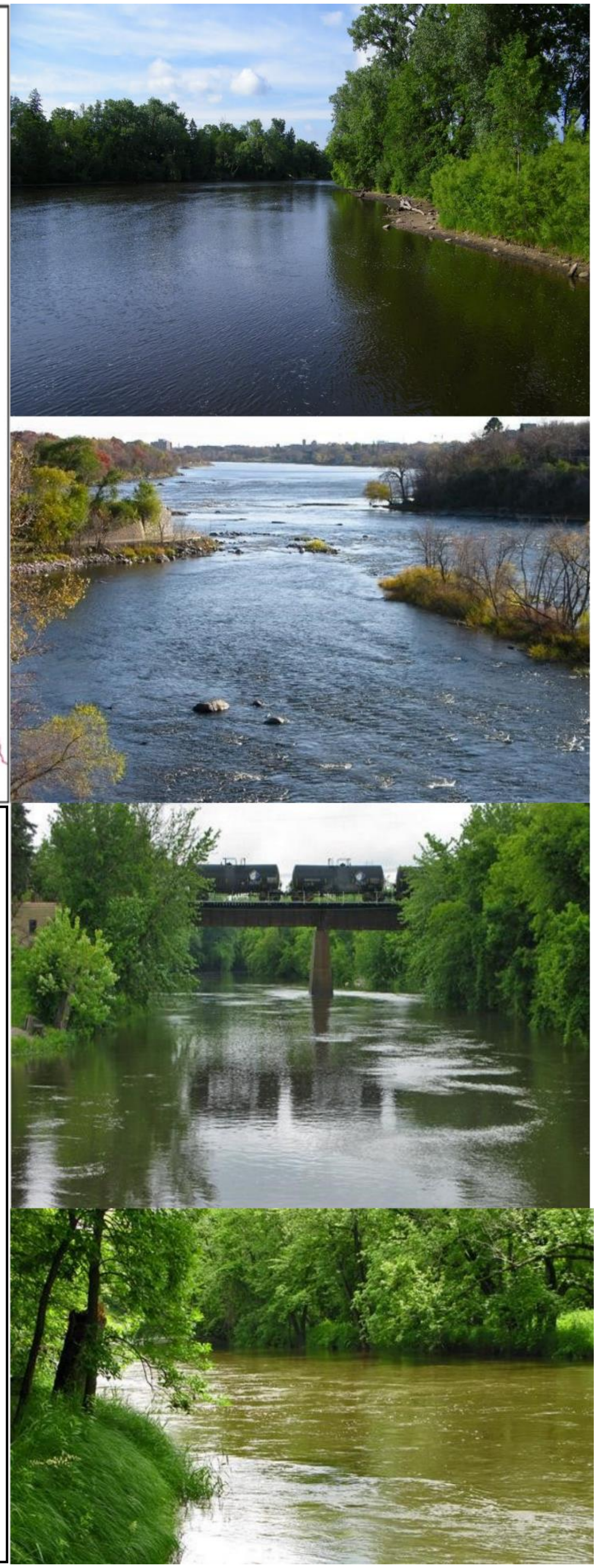
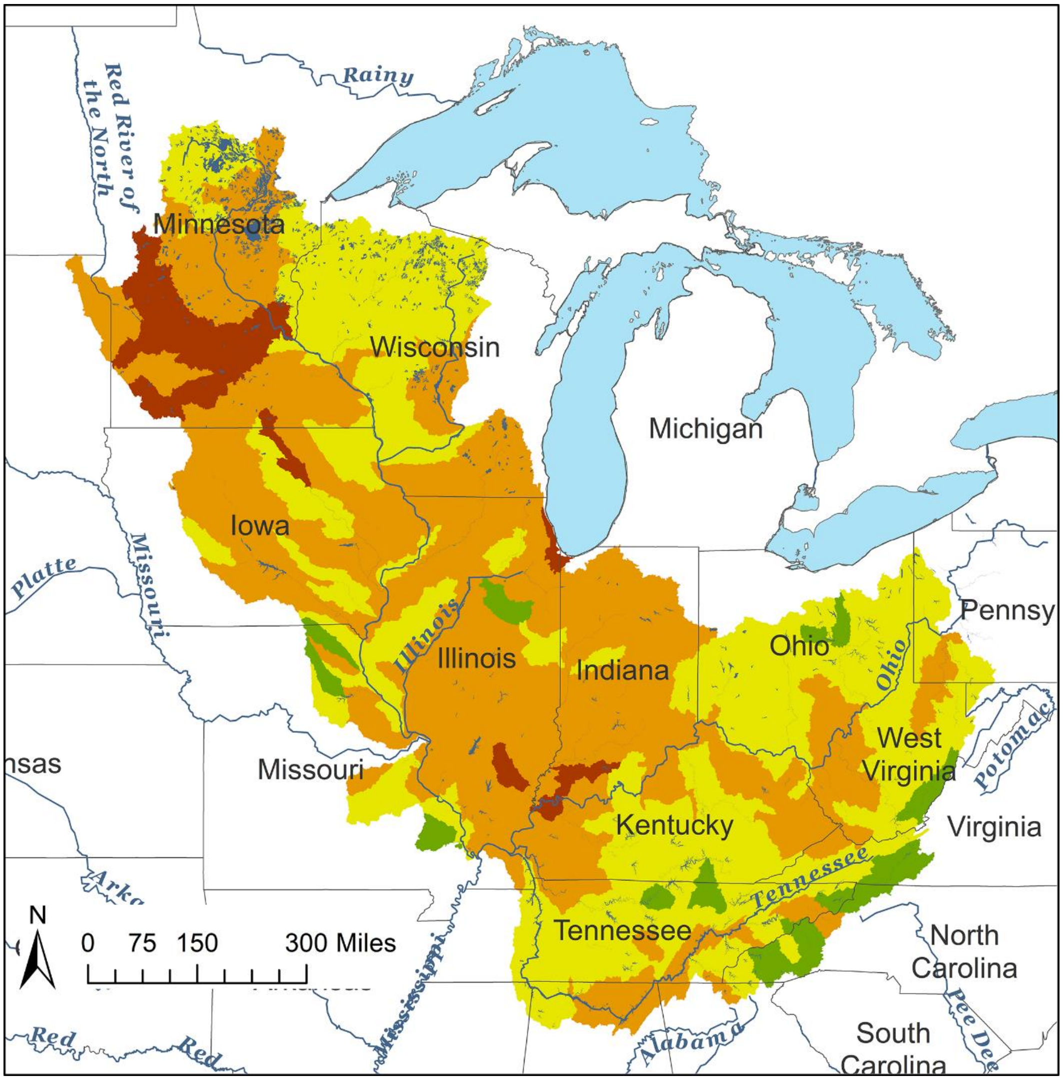


Figure S1.
Above: BCG Levels for major rivers near the Twin Cities, Minnesota. BCG scores shown here are based on fish community data. Authors' knowledge of common recreational uses for these rivers, supplemented with state reports of watershed condition (i.e., including information about water quality likely to affect use such as algal and bacteria growth, turbidity, etc.), were used to overlay uses onto BCG levels more generally across all streams and rivers.

Right panel: From top to bottom, images of the St. Croix River, the Mississippi River at St. Cloud, North Fork of the Crow River, and South Fork of the Crow River. Photo credit for all river photos: Minnesota Pollution Control Agency.



Level 1 Natural State	0%
Level 2 Close to Natural State	4.36%
Level 3 Some Changes Noticeable	41.63%
Level 4 Many Changes Noticeable	48.97%
Level 5 Major Degradation	4.53%
Level 6 Extreme Degradation	0%
Average Level	3.52

Figure S2. Baseline BCG levels in study region

Policy Summary

Description of policy region: Your **local** watershed.
Size of policy region: 17,000 square miles.

	No policy (current conditions)	**Proposed policy (improved conditions)**
Description of change	None	Areas currently at Level 3 improved to Level 2
Water quality near your home	Level 4 - many changes noticeable	Level 4 - many changes noticeable
Water quality throughout region (average)	2.93	2.36
Increase in taxes to your household (per year, for the next 5 years)	None	$150

Figure S3. Example tabular summary of policy scenario

Table S1. Information about biological indices used by each state, and availability of biological monitoring data. BCG= Biological Condition Gradient. NA = no publicly accessible dataset available, data was requested directly from state agency personnel.

State	***Index name***	***Primary agency responsible for data collection/biological index development***	***Publicly available dataset location, if applicable***	***Citations for index development and scoring methodology***
IA	Macroinvertebrate Index off Biotic Integrity (MIBI)	Iowa Department of Natural Resources	https://programs.iowadnr.gov/bionet/; Accessed Oct 03 2019	IDNR 2013. Methodology for Iowa's 2012 water quality assessment, listing, and reporting pursuant to Sections 305(b) and 303(d) of the federal Clean Water Act. Attachment 2: Guidelines for determining Section 305(b) aquatic life use support (ALUS) using stream biocriteria sampling data. Iowa Department of Natural Resources, Environmental Protection Division, Geological Survey and Land Quality Bureau, Watershed Monitoring and Assessment Section; Wilton T.F. 2004. Biological assessment of Iowa's wadeable streams. Project Report. Iowa Department of Natural Resources, Environmental Protection Division, TMDL and Water Quality Assessment Section. Des Moines, Iowa. https://www.iowadnr.gov/Portals/idnr/uploads/watermonitoring/biological/IA_Stream_Bioassessment.pdf
IL	BCG	Illinois Environmental Protection Agency and Illinois Department of Natural Resources	NA; requested and received BCG scores directly from state agency, Jan 4, 2017	Gerritsen and Stamp 2016. Calibration of the Biological Condition Gradient (BCG) for Fish and Benthic Macroinvertebrate Assemblages in Illinois Streams. Prepared for US EPA Region 5 and Illinois Environmental Protection Agency; ILEPA, 2014. Illinois Water Monitoring Strategy 2015-2020. Illinois Environmental Protection Agency, Bureau of Water, Springfield, Illinois. https://www2.illinois.gov/epa/Documents/epa.state.il.us/water/water-quality/monitoring-strategy/monitoring-strategy-2015-2020.pdf
IN	BCG	Indiana Department of Environmental Management	NA; requested and received BCG scores directly from state agency, Jan 4, 2017	Jessup, Stamp and Gerritsen, 2017. Calibration of the Biological Condition Gradient (BCG) for Benthic Macroinvertebrate Assemblages in Indiana Streams. Prepared for US EPA Region 5 and Indiana Department of

				Environmental Management. https://www.in.gov/idem/cleanwater/files/swq_benthic_macro_bcg.pdf
KY	Kentucky Macroinvertebrate Bioassessment Index	Kentucky Department for Environmental Protection	NA; requested and received index scores directly from state agency, Nov 26 2018	KYDEP, 2003. The Kentucky Macroinvertebrate Bioassessment Index. Kentucky Department of Environmental Protection, Division of Water, Water Quality Branch.
MN	BCG	Minnesota Pollution Control Agency	NA; requested and received BCG scores directly from state agency, Jan 9, 2015	Gerritsen, J., Bouchard, R.W., Zheng, L., Leppo, E.W. and C.O. Yoder. Calibration of the biological condition gradient in Minnesota streams: a quantitative expert-based decision system. Freshwater Science 36: 427-451. https://www.journals.uchicago.edu/doi/10.1086/691712
MO	Macroinvertebrate Stream Condition Index (MSCI)	Missouri Department of Natural Resources	https://apps5.mo.gov/mocwis_public/wqa/waterbodySearch.do; Accessed 10-30-2018	MODNR, 2019. Semi-quantitative macroinvertebrate stream bioassessment. Missouri Department of Natural Resources Division of Environmental Quality Environmental Services Program Project Procedure https://dnr.mo.gov/document-search/semi-quantitative-macroinvertebrate-stream-bioassessment-project-procedure Rabeni, C.F., R.J. Sarver, N. Wang, G.S. Wallace, M. Weiland, and J.T. Peterson. 1997. Biological Criteria for Streams of Missouri. Missouri Cooperative Fish and Wildlife Research Unit, University of Missouri, Columbia, MO. 261 pp.
NC	North Carolina Biotic Index, EPT Richness	North Carolina Department of Environmental Quality	https://deq.nc.gov/about/divisions/water-resources/water-resources-science-and-data/water-sciences-home-page/biological-assessment-branch/benthic-macroinvertebrate-assessment-data; Accessed 10-30-2018	NCDEQ, 2016. Standard operating procedures for the collection and analysis of benthic macroinvertebrates. North Carolina Department of Environmental Quality. NCDEQ 2017. Benthic Macroinvertebrate community assessment program (BMCAP) Quality Assurance Project Plan. North Carolina Department of Environmental Quality. 45 pp.
OH	Invertebrate Community Index (ICI)	Ohio Environmental Protection Agency	NA; requested and received BCG scores directly from state agency, Nov 15, 2016	OHEPA, 2015. Biological Criteria for the Protection of Aquatic Life. Volume III. Standardized Biological Field Sampling and Laboratory Methods for Assessing Fish and Macroinvertebrate Communities. Ohio Environmental Protection Agency, Division of Surface Water, Ecological Assessment Section, Ohio EPA Technical Report

				EAS/2015-06-01. https://epa.ohio.gov/static/Portals/35/documents/BioCrit15_Vol3.pdf
TN	Biorecon Index, Semi-Quantitative Index*	Tennessee Department of Environment and Conservation	NA; requested and received index scores directly from state agency, Oct 30, 2018	TNDEC, 2017. Quality System Standard Operating Procedure for Macroinvertebrate Stream Surveys. Tennessee Department of Environment and Conservation Division of Water Resources. 266 pp.
VA	Virginia Stream Condition Index (VASCI)	Virginia Department of Environmental Quality	https://www.deq.virginia.gov/water/water-quality/monitoring/probabilistic-monitoring; requested and received index scores directly from state agency, Nov 9 2018	Tetra Tech, 2003. A stream condition index for Virginia non-coastal streams. 163pp.
WI	Wisconsin Department of Natural Resources	Wisconsin Department of Natural Resources	NA; requested and received BCG scores directly from state agency, Jan 30, 2019	Weigel, B. M. 2003. Development of stream macroinvertebrate models that predict watershed and local stressors in Wisconsin. Journal of the North American Benthological Society 22:123–142. Weigel, B.M. and Dimick, J.J. 2011. Development, validation, and application of a macroinvertebrate-based Index of Biotic Integrity for nonwadeable rivers of Wisconsin. Journal of the North American Benthological Society 30: 665-679.
WV	West Virginia Stream Condition Index (WVSCI)	West Virginia Department of Environmental Protection	NA; requested and received index scores directly from state agency, Nov 28, 2016	Tetra Tech, 2000. A stream condition index for West Virginia wadeable streams. 80 pp. https://dep.wv.gov/WWE/watershed/bio_fish/Documents/WVSCI.pdf

*Tennessee uses two biotic indices: a 'biorecon' index to screen sites, and a semi-quantitative index to make final stream index impairments. We based the BCG proxy for TN on the biorecon index because it was associated with more detailed narrative categories that were used to correspond to BCG levels.

Table S2. Biological index scores assigned to 'BCG proxy' levels in each state without an official BCG.

	Streams		**Rivers**	
State	***Narrative categories and biological index scores used in state biological assessment***	***BCG proxy level assigned to each narrative category and/or index score***	***Narrative categories and biological index scores used in state biological assessment***	***BCG proxy level assigned to each narrative category and/or index score***
IA	Excellent (>75) Good (56-75) Fair (31-55) Poor (<=30)	Level 2 Level 3 Level 4 Level 5 (20-30) Level 6 (<20)		
KY	Excellent Good Fair Poor Very Poor	Level 2 Level 3 Level 4 Level 5 Level 6	Excellent Good Fair Poor Very Poor	Level 2 Level 3 Level 4 Level 5 Level 6
MO	Fully supporting (16-20) Partially supporting (10-14) Non supporting (4-8)	Level 2 (20) Level 3 (16-18) Level 4 (10-14) Level 5 (4-8) Level 6 (<4)	Fully supporting (16-20) Partially supporting (10-14) Non supporting (4-8)	Level 2 (20) Level 3 (16-18) Level 4 (10-14) Level 5 (4-8) Level 6 (<4)
NC	Natural Excellent Good Good-Fair Moderate Fair Poor Severe	Level 1 Level 2 Level 3 Level 4 Level 4 Level 4 Level 5 Level 6	Natural Excellent Good Good-Fair Moderate Fair Poor Severe	Level 1 Level 2 Level 3 Level 4 Level 4 Level 5 Level 6 Level 6
TN	Diverse benthic community (11-15) Ambiguous (7-9) Stressed benthic community (<7)	Level 2 (14-15) Level 3 (11-13) Level 4 (7-9) Level 5 (5-6) Level 6 (<5)	Diverse benthic community (11-15) Ambiguous (7-9) Stressed benthic community (<7)	Level 2 (14-15) Level 3 (11-13) Level 4 (7-9) Level 5 (5-6) Level 6 (<5)
VA	Excellent Good Moderately stressed Severely stressed	Level 2 Level 3 Level 4 Level 5 (21-42) Level 6 (<20)	Excellent Good Moderately stressed Severely stressed	Level 2 Level 3 Level 4 Level 5 (21-42) Level 6 (<20)

Table S3. Variable descriptions

Variable	Description
ASC	Alternative specific constant that equals 1 for the proposed policy; =0 for current policy
BCG for local sub-watershed	BCG score for sub-watershed where the respondent lives; =0 for non-local voting scenarios
BCG for local watershed	BCG score for the respondent's home watershed; =0 if policy involved a different spatial unit
BCG for local 3-watershed group	BCG score for a local group of three watersheds; =0 if policy involved a different spatial unit
BCG for full study area	BCG score for the entire study area; =0 if policy involved a different spatial unit
BCG for non-local watershed	BCG score for a non-local watershed; =0 if policy involved a different spatial unit
BCG for non-local 3-watershed group	BCG score for the local group of three watersheds; =0 if policy involved a different spatial unit
Cost	Cost of the policy, an annual tax payable over five years; =0 for "no policy"
Percent in-state, local	The percentage of the impacted area contained within the respondent's home state; =0 for a non-local policy
Percent in-state, non-local	The percentage of the impacted area contained within the respondent's home state; =0 for a local policy

Table S4. Model 1, all voting scenarios

	Means	Standard deviations
ASC	0.180***	1.58***
	(0.0689)	(0.0505)
BCG for local sub-watershed	–0.735***	0.975***
	(0.0764)	(0.0724)
BCG for local watershed	–0.990***	0.639***
	(0.112)	(0.215)
BCG for local 3-watershed group	–0.909***	0.0489
	(0.108)	(0.172)
BCG for full study area	–0.898***	0.156
	(0.109)	(0.167)
BCG for non-local watershed	–0.817***	0.650***
	(0.0739)	(0.178)
BCG for non-local 3-watershed group	–0.943***	1.25***
	(0.0864)	(0.160)
Cost	–0.00604***	
	(0.000211)	
Individuals	2,000	
Individuals × voting scenarios	17,653	
Log-Likelihood at solution	–9,455.9998	
McFadden's R^2	0.2272	

Notes: Cluster-robust standard errors (clustered by individual) are in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table S5. Model 2, local and non-local watershed voting scenarios

	Means	Standard deviations
ASC	–0.0491	1.57***
	(0.111)	(0.0828)
ASC × Percent in-state, local	0.00682***	
	(0.00200)	
ASC × Percent in-state, non-local	0.0175***	
	(0.00440)	
BCG for local sub-watershed	–0.641***	0.622**
	(0.138)	(0.270)
BCG for local watershed	–0.959***	0.832**
	(0.182)	(0.351)
BCG for non-local watershed	–1.04***	1.08***
	(0.113)	(0.174)
Cost	–0.00627***	
	(0.000338)	
Individuals	1,730	
Individuals × voting scenarios	7,027	
Log-Likelihood at solution	–3923.9483	
McFadden's R^2	0.1944	

Notes: Cluster-robust standard errors (clustered by individual) are in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table S6. Descriptive statistics for selected survey questions and socio-economic characteristics

Variable	Obs	Mean	Std. Dev.	Min	Max
Voted as if Would Pay					
Disagree	2000	.026	.161	0	1
Neutral	2000	.149	.357	0	1
Agree	2000	.824	.381	0	1
Voted as if Outcome Achieved					
Disagree	2000	.031	.173	0	1
Neutral	2000	.168	.374	0	1
Agree	2000	.801	.399	0	1
Voted as if Informed Policy					
Disagree	2000	.033	.177	0	1
Neutral	2000	.207	.405	0	1
Agree	2000	.76	.427	0	1
Importance of Policy Area Size					
Little or no effect	2000	.246	.431	0	1
Moderate effect	2000	.458	.498	0	1
Large effect	2000	.296	.457	0	1
Importance of Water Quality Change					
Little or no effect	2000	.069	.253	0	1
Moderate effect	2000	.307	.461	0	1
Large effect	2000	.625	.484	0	1
Importance of Policy Cost					
Little or no effect	2000	.116	.32	0	1
Moderate effect	2000	.357	.479	0	1
Large effect	2000	.527	.499	0	1
Recreation (=1 if trips in last year)	2000	.706	.456	0	1
Farthest Trip Distance (in last year)					
Less than 25 miles	1408	.305	.46	0	1
25 to 49 miles	1408	.173	.379	0	1
50 to 99 miles	1408	.136	.343	0	1
100 to 149 miles	1408	.103	.304	0	1
150 to 500 miles	1408	.181	.385	0	1
More than 500 miles	1408	.102	.303	0	1
Female (=1 if female)	2000	.529	.499	0	1
Age (in years)	2000	52.905	16.602	18	93
Race/Ethnicity					
White, non-Hispanic	2000	.82	.385	0	1
Black, non-Hispanic	2000	.081	.272	0	1
Asian, non-Hispanic	2000	.018	.133	0	1
Other, non-Hispanic	2000	.01	.1	0	1
2 or more races, non-Hispanic	2000	.026	.161	0	1
Hispanic	2000	.045	.208	0	1
Education					
Less than HS	2000	.02	.14	0	1
HS graduate or equivalent	2000	.169	.374	0	1
Vocational/tech school	2000	.366	.482	0	1
Bachelor's degree	2000	.256	.436	0	1
Post grad study/professional degree	2000	.19	.392	0	1
Married (=1 if married)	2000	.588	.492	0	1
Retired (=1 if retired)	2000	.238	.426	0	1
Household Income (in $1000s)	2000	77.285	56.681	2.5	250
Metro (=1 if lives in metro area)	2000	.743	.437	0	1
Household Size (number of persons)	2000	2.744	1.421	1	6

Table S7. Effects of socio-economic characteristics on the willingness-to-pay for water quality improvements

	Marginal WTP ($)	Std. Err. of WTP
Female (=1 if female)	17.78	14.68
Age (in years)	–1.54***	0.58
Race/Ethnicity		
White, non-Hispanic	<reference category>	
Black, non-Hispanic	19.91	27.88
Asian, non-Hispanic	10.75	43.08
Other, non-Hispanic	–19.10	90.50
2 or more races, non-Hispanic	44.71	49.33
Hispanic	–33.39	34.80
Education		
Less than HS	<reference category>	
HS graduate or equivalent	85.59*	51.82
Vocational/tech school	90.96*	50.32
Bachelor's degree	124.45**	51.44
Post grad study/professional degree	134.81**	52.75
Married (=1 if married)	–5.07	16.17
Retired (=1 if retired)	24.62	21.02
Household Income (in $1000s)	0.72***	0.14
Metro (=1 if lives in metro area)	14.17	16.78
Household Size (number of persons)	5.49	5.92

Notes: Cluster-robust standard errors (clustered by individual) are in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

SI References

1. J. Finlay, Stream size and human influences on ecosystem production in river networks. *Ecosphere* 2, art87.
2. MPCA, *North Fork Crow River: Watershed assessment and trends update*. WQ-WS3-07010204c, Minnesota Pollution Control Agency, St. Paul, MN (2021). https://www.pca.state.mn.us/sites/default/files/wq-ws3-07010204c.pdf
3. MPCA, *South Fork Crow River Watershed Monitoring and Assessment Report*. WQ-WS3-07010205b, Minnesota Pollution Control Agency, St. Paul, MN (2016). https://www.pca.state.mn.us/sites/default/files/wq-ws3-07010205b.pdf
4. U.S. Environmental Protection Agency, *A Practitioner's Guide to the Biological Condition Gradient: A Framework to Describe Incremental Change in Aquatic Ecosystems*. EPA-842-R-16-001, U.S. Environmental Protection Agency, Washington, D.C. (2016). https://www.epa.gov/sites/default/files/2016-02/documents/bcg-practioners-guide-report.pdf
5. R. W. Bouchard, *Development of Biological Criteria for Tiered Aquatic Life Uses*. Minnesota Pollution Control Agency, St. Paul, Minnesota (2016). https://www.pca.state.mn.us/sites/default/files/wq-bsm4-02.pdf
6. J. Gerritsen, J. Stamp, *Calibration of the Biological Condition Gradient (BCG) for fish and benthic macroinvertebrate assemblages in Illinois streams*. Illinois Environmental Protection Agency, Springfield, Illinois (2016).
7. D. McFadden, "Conditional logit analysis of qualitative choice behavior" in Frontiers of Econometrics, P. Zarembka, Ed. (Academic Press, 1974), pp. 105-142.
8. D. Revelt, K. Train, Mixed logit with repeated choices: Households' choices of appliance efficiency level. *Rev. Econ. Stat.* **80**(4), 647-657 (1998).
9. W. H. Green, *Econometric Analysis*, 8th Edition (Pearson, 2018).